

Petruş Alexandrescu

Introducere
în
statistica socială

CUPRINS

Prefață

Capitolul I

Elemente de teoria probabilităților

Capitolul II

Elemente de statistică descriptivă

Capitolul III

Variabile aleatoare. Proprietăți. Caracteristici.

Capitolul IV

Indicatori ai caracteristicilor cantitative

Capitolul V

Corelația rangurilor.

Capitolul VI

Analiza de regresie

Capitolul VII

Analiza de dependenta

Capitolul IX

Chestionarul de opinie. Elemente privind proiectarea chestionarului.

Capitolul X

Teste de semnificație

Capitolul XI

Modele de analiză a caracteristicilor calitative

Capitolul XII

O metodă de analiză scalară și ierarhizare

Capitolul XIII

Sisteme electorale

Capitolul XIV

Modele de subiecte de examen

Bibliografie selectivă

Prefață

Lucrarea de față urmează în esență cursul de statistică socială de un semestru, predat de autor la anul II al Facultății de Sociologie și Psihologie a Universității Spiru Haret. Pentru a fi utilă în primul rând studenților acestei facultăți, lucrarea urmărește pe de o parte să familiarizeze cititorul cu elementele de statistică matematică necesare în abordarea și înțelegerea unui fenomen social. Pentru aceasta, am ținut cont de faptul că numeroși studenți ai acestei facultăți au formație umanistă încă din liceu. Acest lucru a făcut ca interesul lor pentru disciplinele realiste să fie scăzut. Întâlnirea acestora cu statistica în cadrul facultății, este privită cu o anumită rețineră. Rolul profesorului în acest caz este, de a face, pe cât posibil, un curs foarte accesibil, atractiv, și care să stârnească curiozitatea studentului (măcar al celui care își cunoaște interesul și știe de ce a venit la această facultate). Cursul predat, a încercat permanent să țină seama de acest deziderat iar cursul scris încearcă să-l urmeze îndeaproape.

Dar, pe lângă a fi accesibil și atractiv, cursul trebuie să fie util. Odată cu înțelegerea rolului statisticii în realitatea socială, este important să se înțeleagă metodele, tehnicile sale, dar mai ales este important să se înțeleagă **gândirea statistică**. Nu

trebuie să utilizăm o metodă sau alta pentru că am auzit de ea sau pentru că utilizarea unor metode statistice sonore ne-ar scoate din impas sau ne-ar pune în situația comodă de a ne așeza în spatele lor și a ne mulțumi astfel cu orice rezultat obținut.

Utilizarea statisticii în mod adecvat este deosebit de utilă. Dar utilizarea statisticii poate fi și nocivă atunci când se face în mod mecanic, fără a se înțelege utilitatea sa și mai ales când, cum și în ce fel poate fi folosită.

Am încercat să lămurim și aceste lucruri dealungul cursului.

Autorul, de formație matematician, și-a făcut ucenicia în metodologia științelor sociale delungul anilor în cadrul Institutului de Sociologie al Academiei Române. Aici a avut posibilitatea să participe la numeroase cercetări concrete în colective interdisciplinare, să învețe și să experimenteze o serie de metode și tehnici. Anii de după Revoluție au putut fi mult mai profitabili din acest punct de vedere.

Comenzile sociale ne-au ajutat să ținem pasul cu realitatea socială și să găsim soluții practice chiar și atunci când teoria nu ne ajuta. Am învățat din greșelile noastre ca și din ale altora, îmbunătățindu-ne stilul și bagajul de cunoștințe. Toate aceste cunoștințe căpătate le putem împărtăși celor tineri pentru a le netezi drumul și a-i ajuta să devină specialiștii de mâine în

domeniul științelor sociale. Facultatea de Sociologie și Psihologie a Universității “Spiru Haret” mi-a oferit această ocazie.

Prezenta lucrare se adresează în primul rând studenților Facultății de Sociologie și Psihologie dar și tuturor celor care doresc să se inițieze în statistica socială.

Autorul

Capitolul I

ELEMENTE DE TEORIA PROBABILITĂȚILOR

I. Câmp de probabilitate finit

Experimentul statistic este un procedeu care poate fi repetat în condiții similare și în urma căruia se obțin rezultate ce pot fi observate, măsurate și apoi interpretate.

Experimentul statistic are un caracter aleator, în sensul că rezultatul acestuia variază la întâmplare; de aceea îl vom numi adesea, și experiență aleatoare.

Rezultatul unei experiențe aleatoare se va numi **probă**.

Exemplu. Aruncarea unui zar constituie o experiență aleatoare. Rezultatele posibile acestei experiențe sunt concretizate în apariția uneia din fețele: 1, 2, 3, 4, 5, 6.

Definiție. Realizarea sau nerealizarea unei anumite situații, legată de experiența aleatoare avută în vedere, după efectuarea experienței, se numește **eveniment statistic**.

Prin **evenimentul elementar** vom înțelege acel eveniment care poate fi realizat numai de o singură probă. Celelalte evenimente le vom numi **compuse**.

De exemplu, evenimentul de apariție a feței cu numărul 6 este un eveniment elementar. Evenimentul de apariție a unei fețe cu număr par este realizat de una din probele $\{2\}$, $\{4\}$, $\{6\}$.

Evenimentul sigur este evenimentul care se realizează cu certitudine la fiecare efectuare a experienței.

Evenimentul $\{1,2,3,4,5,6\}$ este evenimentul sigur al experienței.

Evenimentul imposibil este evenimentul care nu se realizează la nici o efectuare a experienței.

Evenimentul imposibil se notează prin $\emptyset$.

Două evenimente A și B se numesc contrare dacă nerealizarea unuia este echivalentă cu realizarea celuilalt; asta înseamnă că nu există nici o probă care să le realizeze simultan pe amândouă în schimb, orice probă realizează unul din cele două evenimente.

În termeni de teoria mulțimilor, astfel de evenimente sunt asociate mulțimilor complementare. Astfel, $B = C_A$ iar $A = C_B$.

Evenimentele A și B se numesc compatibile dacă se pot realiza simultan, adică dacă există probe care le realizează în același timp pe A și pe B . În caz contrar, evenimentele A și B se numesc **incompatibile**.

În primul caz, comparând cu mulțimile, avem $A \cap B \neq \emptyset$ iar în al doilea caz, $A \cap B = \emptyset$.

Evenimentul A implică evenimentul B și scriem $A \subset B$, dacă realizarea lui A implică realizarea lui B .

Operații cu evenimente

Dacă A și B sunt două evenimente, numim reuniunea lor și notăm $A \cup B$, evenimentul a cărui realizare constă în realizarea a cel puțin unuia din cele două evenimente.

În mod asemănător, $A \cap B$ este evenimentul care se realizează odată cu realizarea simultană a evenimentelor A și B .

Dacă evenimentele A și B sunt incompatibile atunci $A \cap B = \emptyset$.

Definiția probabilității

Să considerăm experiența de aruncare a unui zar și A evenimentul de apariție a feței cu numărul 5. Să repetăm această experiență de 10 ori. Să presupunem că aruncând cu zarul de 10 ori, de trei ori a apărut fața cu numărul 5. Raportul

$$f = \frac{3}{10}$$

se numește **frecvență de apariție**.

Frecvența de apariție este un număr subunitar $0 \leq f_n \leq 1$.

Două evenimente A și B se numesc **egal posibile** dacă au aceeași șansă de a se realiza.

Dacă la experiența de aruncare a unui zar, A este evenimentul în care apare fața 5 și B este evenimentul de apariție a feței 3, atunci evenimentele A și B sunt egal posibile.

Definiție. Numim probabilitate a unui eveniment, raportul dintre numărul cazurilor egal posibile care realizează evenimentul sau cazurile favorabile și numărul cazurilor egal posibile

Exemple 1) La experiența de aruncare a unei monede, probabilitatea de a apărea stema este:

$$p = \frac{1}{2}$$

2) Care este probabilitatea ca aruncând două zaruri să obținem o dublă, adică (1,1) sau (2,2), ..., sau (6,6)?

$$p = \frac{6}{36} = \frac{1}{6}$$

Proprietăți ale probabilităților.

Dacă $p(A)$ este probabilitatea evenimentului A, atunci au lor proprietățile:

- 1) $0 \leq p(A) \leq 1$
- 2) $p(E) = 1$, unde E este evenimentul sigur
- 3) $p(\emptyset) = 0$, unde $\emptyset$ este evenimentul imposibil
- 4) $p(A \cup B) = p(A) + p(B)$, dacă $A \cap B = \emptyset$

Dacă $A \cap B \neq \emptyset$, atunci proprietatea 4) devine:

$$4') p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$5) p(A) + p(\bar{A}) = 1$$

unde $\bar{A}$ este evenimentul contrar lui A.

Definiție. Dacă A și B sunt două evenimente și dacă

$$p(A \cap B) = p(A) \cdot p(B)$$

atunci spunem că evenimentele A și B sunt **independente**.

Dacă A,B,C sunt trei evenimente, atunci ele sunt independente dacă sunt indeplinite relațiile:

$$p(A \cap B) = p(A) \cdot p(B)$$

$$p(A \cap C) = p(A) \cdot p(C)$$

$$p(B \cap C) = p(B) \cdot p(C)$$

$$p(A \cap B \cap C) = p(A)p(B)p(C)$$

Formule clasice de probabilitate

1. Formule pentru calculul unor probabilități

a) Fie A,B – 2 evenimente. Atunci are loc formula:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

b) Dacă A,B,C sunt 3 evenimente, atunci are loc formula:

$$p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(A \cap B) - p(A \cap C) - p(B \cap C) + p(A \cap B \cap C)$$

Această formulă se poate generaliza la n evenimente și se obține o formulă care poartă denumirea de formula lui H.Poincaré.

Aplicație. O urnă conține 4 bile albe și 6 bile negre iar altă urnă conține 7 bile albe și 3 bile negre. Din fiecare urnă se extrage câte o bilă. Care este probabilitatea ca cel puțin o bilă să fie albă?

Rezolvare. Notăm cu A evenimentul ca bila extrasă din prima urnă să fie albă și cu B evenimentul ca bila extrasă din a doua urnă să fie albă. Vom calcula probabilitatea evenimentului $A \cup B$:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Dar evenimentele A și B sunt independente, rezultă că

$$p(A \cap B) = p(A) \cdot p(B) = \frac{4}{10} \cdot \frac{7}{10} = \frac{28}{100}$$

Urmează că:

$$p(A \cup B) = \frac{4}{10} + \frac{7}{10} - \frac{28}{100} = \frac{110 - 28}{100} = \frac{82}{100} = 0,82$$

2. Scheme clasice de probabilitate

a) Schema lui Poisson

Să presupunem că avem n urne:

$$U_1, U_2, \dots, U_n$$

care conțin bile albe și negre. Dacă p_i este probabilitatea cu care este extrasă o bilă albă din urna U_i , se cere probabilitatea de a extrage k bile albe $0 \leq k \leq n$, atunci când din fiecare urnă se extrage câte o bilă.

Ca regulă, să reținem că determinarea probabilității cerute este similară cu determinarea coeficientului lui x^k din dezvoltarea polinomului

$$P(x) = (p_1x + q_1)(p_2x + q_2) \dots (p_nx + q_n).$$

Vom lămuri acest lucru printr-un exemplu.

Exemplu. Într-un atelier sunt 3 strunguri care execută piese. Primul dă 1,2% rebuturi, al doilea 1,5% și al treilea, 1,4%. Se ia la întâmplare câte o piesă de la fiecare strung. Se cere probabilitatea ca 2 din piese să fie bune și una să fie rebut.

Probabilitatea căutată va fi coeficientul lui x^2 din dezvoltarea polinomului

$$(p_1x + q_1)(p_2x + q_2)(p_3x + q_3)$$

Se constată că: $n = 3$, $k = 2$,

$$q_1 = 0,012, q_2 = 0,015, q_3 = 0,014$$

$$p_1 = 0,988, p_2 = 0,985, p_3 = 0,986$$

Așadar, coeficientul lui x^2 va fi:

$$\begin{aligned} p_1p_2q_3 + p_1p_3q_2 + p_2p_3q_1 &= 0,988 \cdot 0,985 \cdot 0,014 + 0,988 \cdot \\ &\cdot 0,986 \cdot 0,015 + 0,985 \cdot 0,986 \cdot 0,012 = 0,013 + 0,015 + \\ &+ 0,012 = 0,04 \end{aligned}$$

în timp ce probabilitatea ca toate piesele extrase să fie bune este:

$$P = p_1p_2p_3 = 0,988 \cdot 0,985 \cdot 0,986 = 0,959$$

b) Schema lui Bernoulli.

Este un caz particular al schemei Poisson în care cele n urne au conținuturi identice. În acest caz:

$$p_1 = p_2 = \dots = p_n = p \text{ și}$$

$$q_1 = q_2 = \dots = q_n = q = 1 - p$$

Problema este aceeași, anume, de a extrage câte o bilă din fiecare urnă și de a calcula probabilitatea ca din cele n bile extrase, k să fie albe și $n-k$ negre.

Probabilitatea cerută este coeficientul lui x^k din dezvoltarea binomului:

$$P(x) = (px+q)^n$$

adică:

$$C_n^k p^k q^{n-k}$$

Să mai constatăm că problema extragerii a n bile din n urne identice, câte una din fiecare urnă, este similară cu aceea a extragerii succesive a n bile din aceeași urnă, punând de fiecare dată bila extrasă înapoi.

Aplicație. Aruncăm o monedă de 6 ori. Se cere probabilitatea ca stemă să apară o singură dată și banul de 5 ori.

$$\text{Se constată că } n = 6, k = 1, p = q = \frac{1}{2}$$

Atunci

$$P = C_6^1 \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^5 = \frac{6}{2^6} = \frac{3}{32} = 0,094$$

c) Schema bilei neîntoarse

Să presupunem că o urnă conține a bile albe și b bile negre. Din această urnă se extrag n bile, fără a pune bila extrasă înapoi. Se cere probabilitatea ca din cele n bile extrase, α să fie albe și $\beta = n - \alpha$ să fie negre.

Probabilitatea căutată va fi dată de formula:

$$\frac{C_a^\alpha \cdot C_b^\beta}{C_{a+b}^{\alpha+\beta}}$$

unde $n < a + b$, iar $\alpha + \beta = n$

Aplicație. Într-o urnă sunt 100 de bile, 40 roșii și 60 albe. Care este probabilitatea ca extrăgând 4 bile, două să fie albe?

Suntem în situația schemei bilei neîntoarse cu $a = 40$, $b = 60$, $\alpha = 2$, $\beta = 2$.

Probabilitatea căutată va fi:

$$\frac{C_{40}^2 \cdot C_{60}^2}{C_{100}^4} = \frac{\frac{40 \cdot 39}{1 \cdot 2} \cdot \frac{60 \cdot 59}{1 \cdot 2}}{\frac{100 \cdot 99 \cdot 98 \cdot 97}{1 \cdot 2 \cdot 3 \cdot 4}} = 0,015$$

Capitolul II

ELEMENTE DE STATISTICĂ DESCRIPTIVĂ

Statistica socială se ocupă cu gruparea, analiza și interpretarea datelor referitoare la un fenomen social. Totodată, cu mijloacele statisticii sociale se pot efectua o serie de previziuni privind producerea fenomenului în viitor.

Statistica socială sau metodologia statistică pe care o avem în vedere presupune două etape:

- **statistica descriptivă** este un proces de culegere a datelor despre un fenomen social și înregistrarea acestora.
- **statistica matematică** care se ocupă cu gruparea datelor, analiza și interpretarea acestora în vederea explicării fenomenului social și a posibilității efectuării unor predicții asupra derulării fenomenului în viitor.

Populația statistică. Prin conceptul de populație statistică vom înțelege orice mulțime care formează obiectul de studiu al analizei statistice.

Elementele unei populații statistice le numim **unități statistice** (sau indivizi).

O analiză statistică are în vedere anumite caracteristici. Astfel, dacă luăm ca exemplu rezultatele obținute la examenul de statistică a unei colectivități de studenți, atunci putem înregistra notele obținute după:

- caracteristica “sex”: (M, F).
- caracteristica “grupe de vârstă”, etc.

În cazul efectuării unui studiu în care avem în vedere numărul locuitorilor dintr-o anumită zonă, mulțimea localităților din acea zonă poate constitui populația statistică. O caracteristică de studiu ar putea fi numărul locuitorilor din fiecare localitate.

O caracteristică care se poate măsura se va numi **caracteristică cantitativă**.

În cazul exemplului anterior, rezultatul obținut la examen se măsoară în note, deci este o caracteristică cantitativă.

Caracteristica “grupe de vârstă”, “venitul pe familie”, etc pot fi considerate drept caracteristici cantitative.

Caracteristicile care nu pot fi măsurate se numesc **caracteristici calitative**.

Un exemplu de caracteristică calitativă poate fi înregistrarea răspunsurilor la întrebarea: Cum vă place emisiunea X de la T.V?: mult, puțin, deloc.

Caracteristicile calitative sunt cel mai greu de înregistrat într-o analiză statistică.

Există caracteristici care pot lua numai valori întregi. Acestea se va numi caracteristici discrete. Exemplu: numărul de localități dintr-un județ, numărul persoanelor dintr-o gospodărie, etc.

Există și **caracteristici continue** al căror număr de valori este infinit. Un astfel de exemplu îl constituie “vârsta”. Structura acestei caracteristici pe grupe de vârstă o transformă într-o variabilă discretă, după cum se poate vedea și din tabelul următor, în care am efectuat grupări ale populației adulte:

Grupe de vârstă	18-25 ani	26-35 ani	36-45 ani	46-55 ani	56-62 ani	peste 62 ani
-----------------	-----------	-----------	-----------	-----------	-----------	--------------

În analizele statistice suntem nevoiți să facem astfel de grupări pentru a simplifica etapele cercetării.

Frecvențe absolute, frecvențe relative, frecvențe cumulate.

Să considerăm exemplul unei colectivități de 200 de studenți, dintre care 86 studenți și 114 studente. Această structură pe sexe poate fi înregistrată în tabelul:

	Masculin	Feminin	Total
Frecvențe absolute	86	114	200
Frecvențe relative	43%	57%	100%

Frecvența absolută înseamnă numărul de subiecți înregistrați; frecvența relativă este raportarea la total, adică:

$$\frac{86}{200} = 0,43 = 43\%$$

Să presupunem că cei 86 de studenți de sex masculin îi distribuim după rezultatele obținute la un examen astfel:

Nota obținută	< 5	5	6	7	8	9	10	Total
Număr subiecți	6	6	4	14	16	25	15	86
Frecvențe relative	7%	7%	5%	16%	19%	29%	17%	100%
Frecvențe absolute cumulate	6	12	16	30	46	71	86	

Frecvențele cumulate se folosesc mai ales atunci când urmărim evoluția unui fenomen.

Serii statistice. Reprezentări grafice

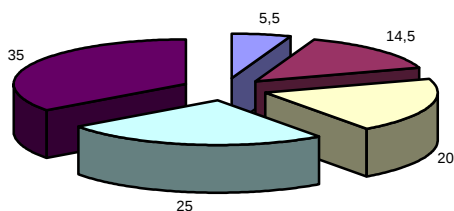
Seriile statistice sunt serii de date care se înșiruiesc după o anumită caracteristică. Dacă această caracteristică este timpul, seriile se numesc temporale.

Să presupunem că într-o întreprindere industrială se fac investiții procentuale în cinci sectoare astfel:

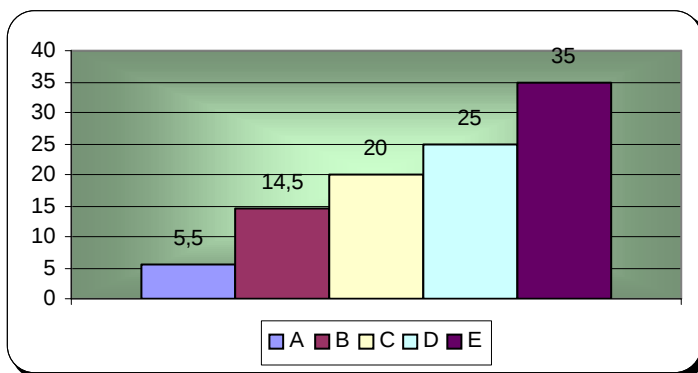
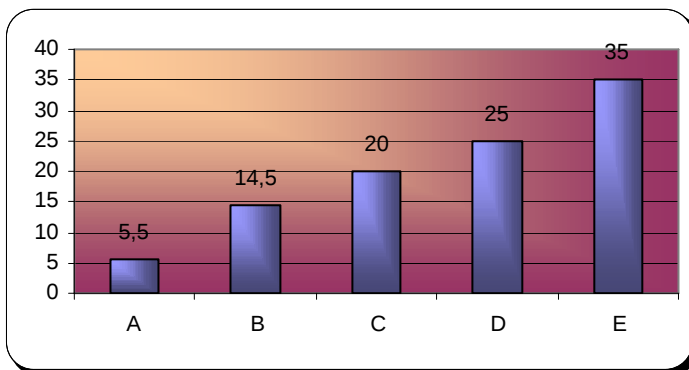
Sectorul procentuale	Investiții
A	5,5%
B	14,5%
C	20%
D	25%
E	35%

Reprezentarea investițiilor se poate face în mai multe feluri.

a) Cu ajutorul diagramei:



b) prin histograme:



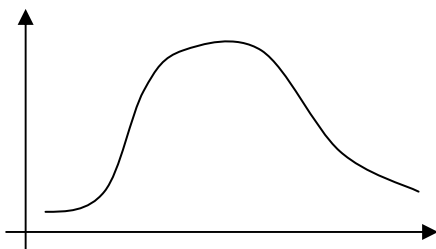
c) cu ajutorul **poligoanelor de frecvență**

Există reprezentări statistice în care apare numai poligonul frecvențelor fără histograme.



- creșterea producției industriale
pe lunile unui an.

d) curba frecvențelor

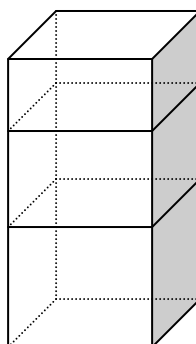


Poligonul frecvențelor unei variabile aleatoare poate fi aproximat cu o curbă – numită **curba frecvențelor** sau **curba de distribuție**.

Alte reprezentări sunt reprezentări gen hărți numite **cartodiagrame**.

Diagrame de structură

40%	60%
Rural	Urban



- populație > 60 ani

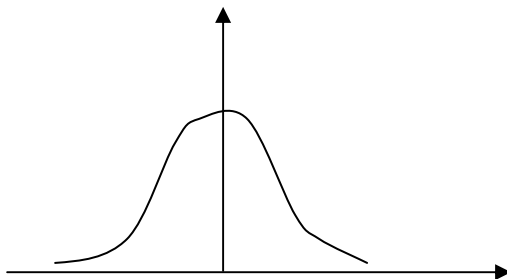
- populație activă

- populație tânără

În privința distribuției curbelor de frecvență, putem vorbi de următoarea clasificare:

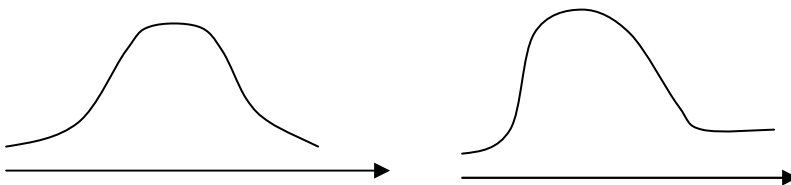
- curbe simetrice (sau normale)
- curbe asimetrice

Distribuția simetrică (sau normală) este distribuția lui Gauss:

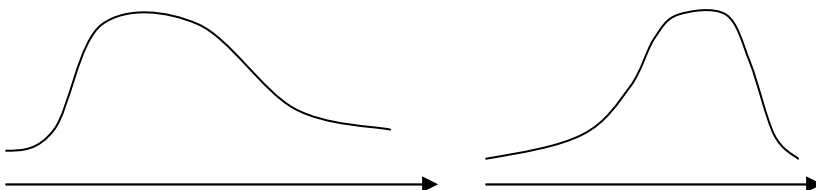


Distribuțiile asimetrice sunt și ele de mai multe feluri:

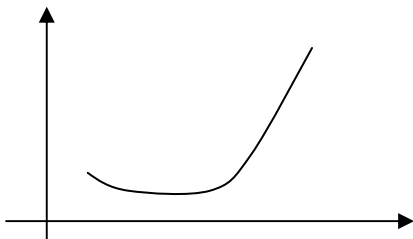
- distribuții ușor asimetrice



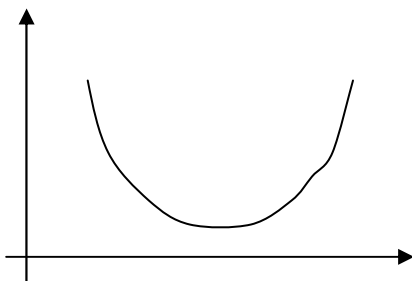
- distribuții pronunțat asimetrice



- distribuții în formă de **J**



- distribuții în formă de **U**



Capitolul III

VARIABLE ALEATOARE.

PROPRIETĂȚI. CARACTERISTICI.

O variabilă aleatoare X este un tabel de forma

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} \quad (1)$$

unde x_i sunt valorile pe care le poate lua variabila cu probabilitățile p_i . În plus, trebuie îndeplinite condițiile:

$$1. 0 \leq p_i \leq 1, \quad i=1, 2, \dots, n$$

$$2. p_1 + p_2 + \dots + p_n = 1$$

Spunem că tabelul (1) reprezintă distribuția sau repartiția variabilei aleatoare X .

Pot exista variabile aleatoare diferite cu aceeași distribuție.

Să considerăm experiența aruncării unui zar. Întru-cât fiecare față are aceleași șanse de apariție în raport cu celelalte, probabilitatea de apariție a oricărei fețe va fi: $p = \frac{1}{6}$.

Variabila aleatoare asociată acestei experiențe are distribuția:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Operații cu variabile aleatoare

Considerăm variabila aleatoare X de repartiție:

$$X \quad \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

cu valorile $x_1, x_2, \dots, x_n$ care sunt luate cu probabilitățile $p_1, p_2, \dots, p_n$.

Putem scrie acest lucru astfel:

$$P(X = x_1) = p_1, P(X = x_2) = p_2, \dots \text{ etc.}$$

Dacă a este o constantă nenulă, atunci putem vorbi de suma variabilei aleatoare X cu constanta a și produsul cu constanta a . Astfel, vom obține noi variabile aleatoare: $a + X$ și aX de distribuții:

$$a + X \quad \begin{pmatrix} a + x_1 & a + x_2 & \dots & a + x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

și

$$aX \quad \begin{pmatrix} ax_1 & ax_2 & \dots & ax_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Dacă avem variabilele aleatoare

$$X \quad \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$Y \quad \begin{pmatrix} y_1 & y_2 & \dots & y_m \\ q_1 & q_2 & \dots & q_m \end{pmatrix}$$

atunci putem defini variabilele $X + Y$ și XY astfel:

$$X + Y \quad \begin{pmatrix} x_1 + y_1 & x_1 + y_2 & \dots & x_1 + y_m & \dots & x_n + y_m \\ p_{11} & p_{12} & \dots & p_{1m} & \dots & p_{nm} \end{pmatrix}$$

astfel încât

$$\sum_{j=1}^m \sum_{i=1}^n p_{ij} = 1$$

iar p_{ij} este probabilitatea realizării simultane a evenimentelor $(X = x_i)$ și $(Y = y_j)$. Dacă evenimentele $(X = x_i)$ și $(Y = y_j)$ sunt independente, atunci

$$p_{ij} = P(X = x_i \text{ și } Y = y_j) = P((X = x_i) \cap (Y = y_j)) = \\ P(X = x_i) \cdot P(Y = y_j) = p_i \cdot q_j$$

Exemplu: Fie variabilele aleatoare X și Y de repartiții:

$$X \quad \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

și

$$Y \quad \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}$$

Atunci

$$X+Y \quad \begin{pmatrix} -1+1 & -1+2 & -1+3 & 0+1 & 0+2 & 0+3 & 1+1 & 1+2 & 1+3 \\ \frac{1}{2} \cdot \frac{1}{3} & \frac{1}{2} \cdot \frac{1}{2} & \frac{1}{2} \cdot \frac{1}{6} & \frac{1}{4} \cdot \frac{1}{3} & \frac{1}{4} \cdot \frac{1}{2} & \frac{1}{4} \cdot \frac{1}{6} & \frac{1}{4} \cdot \frac{1}{3} & \frac{1}{4} \cdot \frac{1}{2} & \frac{1}{4} \cdot \frac{1}{6} \end{pmatrix}$$

sau:

$$X+Y \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{6} & \frac{1}{4} + \frac{1}{12} & \frac{1}{12} + \frac{1}{8} + \frac{1}{12} & \frac{1}{24} + \frac{1}{8} & \frac{1}{24} \end{pmatrix}$$

$$X+Y \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{6} & \frac{1}{3} & \frac{7}{24} & \frac{1}{6} & \frac{1}{24} \end{pmatrix}$$

În mod analog se definește variabila produs.

$$X \cdot Y \begin{pmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_m & \dots & x_n y_m \\ p_{11} & p_{12} & \dots & p_{1m} & \dots & p_{nm} \end{pmatrix}$$

În cazul exemplului de mai sus, variabila $X \cdot Y$ va avea distribuția:

$$X \cdot Y \begin{pmatrix} -1 & -2 & -3 & 0 & 0 & 0 & 1 & 2 & 3 \\ \frac{1}{6} & \frac{1}{4} & \frac{1}{12} & \frac{1}{12} & \frac{1}{8} & \frac{1}{24} & \frac{1}{12} & \frac{1}{8} & \frac{1}{24} \end{pmatrix}$$

sau:

$$X \cdot Y \begin{pmatrix} -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ \frac{1}{24} & \frac{1}{4} & \frac{1}{6} & \frac{1}{4} & \frac{1}{12} & \frac{1}{8} & \frac{1}{24} \end{pmatrix}$$

Să considerăm un alt exemplu. Să presupunem că p este probabilitatea extragerii unei bile albe dintr-o urnă. După

efectuarea primei extrageri și întoarcerea bilei în urnă, repetăm experiența.

Fie X_1 și X_2 variabilele aleatoare asociate celor două experiențe, cu distribuțiile:

$$X_1 \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix} \qquad X_2 \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$$

în care am notat cu 1 valoarea variabilei la apariția bilei albe și cu 0 neapariția unei bile albe.

Atunci

$$X_1+X_2 \begin{pmatrix} 1+1 & 1+0 & 0+1 & 0+0 \\ p^2 & pq & qp & q^2 \end{pmatrix}$$

sau

$$X_1+X_2 \begin{pmatrix} 2 & 1 & 0 \\ p^2 & 2pq & q^2 \end{pmatrix}$$

iar variabila produs:

$$X_1 \cdot X_2 \begin{pmatrix} 1 \cdot 1 & 1 \cdot 0 & 0 \cdot 1 & 0 \cdot 0 \\ p^2 & pq & qp & q^2 \end{pmatrix}$$

adică

$$X_1 \cdot X_2 \begin{pmatrix} 1 & 0 \\ p^2 & 2pq+q^2 \end{pmatrix}$$

Caracteristici ale variabilelor aleatoare

Considerăm variabila aleatoare X de distribuție

$$X \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

cu $p_1 + p_2 + \dots + p_n = 1$

Numim valoare medie a variabilei aleatoare X expresia:

$$M(X) = p_1 x_1 + p_2 x_2 + \dots + p_n x_n = \sum_{i=1}^n p_i x_i$$

Proprietăți

1) $M(a) = a$, unde a este o constantă. Acest lucru înseamnă că valoarea medie a unei constante este aceea constantă.

2) $M(a + X) = a + M(X)$

3) $M(a \cdot X) = a \cdot M(X)$

4) $\alpha \leq M(X) \leq \beta$, unde am notat

$\alpha = \min(x_1, x_2, \dots, x_n)$ adică cea mai mică valoare a lui X

iar $\beta = \max(x_1, x_2, \dots, x_n)$ este cea mai mare valoare a lui X .

5) $M(X+Y) = M(X) + M(Y)$

6) $M(X \cdot Y) = M(X) \cdot M(Y)$

relația ultimă are loc numai dacă variabilele X și Y sunt independente.

Spunem că variabilele X și Y sunt **independente** dacă evenimentele $(X = x_i)$ și $(Y = y_j)$ sunt independente pentru toate cuplurile i și j , $i \in \{1, 2, \dots, m\}$ și $j \in \{1, 2, \dots, n\}$.

Momente

Notăm prin X^k variabila aleatoare cu distribuția

$$\left(\begin{array}{cccc} x_1^k & x_2^k & \dots & x_n^k \\ p_1 & p_2 & \dots & p_n \end{array} \right) \quad (k \in \mathbb{R})$$

Numim **moment de ordinul k** al variabilei X , valoarea medie a variabilei X^k :

$$M_k(X) = M(X^k) = \sum_{i=1}^n p_i x_i^k$$

Variabila $X - M(X)$ se numește **abaterea de la medie** a variabilei X .

Se constată că media acestei variabile aleatoare este 0, deoarece:

$$M(X - M(X)) = M(X) - M(X) = 0$$

Împrăștierea variabilei X se caracterizează prin variabila $|X - M(X)|$ - numită **abatere medie**, cu distribuția:

$$\left(\begin{array}{cccc} |x_1 - m| & |x_2 - m| & \dots & |x_n - m| \\ p_1 & p_2 & \dots & p_n \end{array} \right)$$

unde am notat $m = M(X)$.

Cel mai comod indicator al împrăstierii este dat de expresia

$$M[(X - m)^2]$$

care este un moment centrat de ordinul al doilea. Acesta se mai numește adesea **dispersia** variabilei X și se notează cu σ^2 sau $D(X)$. Avem:

$$\sigma^2 = D(X) = M[(X - m)^2].$$

unde $m = M(X)$.

Proprietăți

- a) $D(a) = 0$, unde a este o constantă, adică dispersia unei constante este 0.
- b) $D(a + X) = D(X)$
- c) $D(aX) = a^2 D(X)$
- d) $D(X+Y) = D(X) + D(Y)$ dacă variabilele X și Y sunt independente. În caz contrar,

$$D(X+Y) = D(X) + D(Y) + 2M[(x - \bar{x})(y - \bar{y})].$$

Variabila

$$\sigma = \sqrt{D(X)} = \sqrt{M(X^2) - M^2(X)}$$

se numește **abatere medie pătratică**.

Inegalitatea lui Cebâșev.

Următoarea inegalitate spune că probabilitatea ca variabila $|X - m|$ să fie mai mică decât o cantitate oricât de mică, ε , este

mai mare sau egală cu expresia $1 - \frac{\sigma^2}{\varepsilon^2}$ unde σ este abaterea medie pătratică a v.a. X iar ε este un coeficient de toleranță sau prag de semnificație.

Inegalitatea lui Cebâșev se va scrie:

$$P(|x - m| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$

Exemplu. Fie X o variabilă aleatoare. Să determinăm dispersia acesteia știind că

$$P(|x - m| < 8) \geq \frac{15}{16}$$

Se constată că $\varepsilon = 8$. Atunci $1 - \frac{\sigma^2}{\varepsilon^2} = \frac{15}{16}$

De unde:

$$\frac{\sigma^2}{\varepsilon^2} = 1 - \frac{15}{16} = \frac{1}{16}$$

De aici rezultă:

$$D(X) = \sigma^2 = \frac{\varepsilon^2}{16} = \frac{64}{16} = 4$$

Dacă X și Y sunt două variabile aleatoare, notăm prin

$$\mu_{xy} = M[(X - \bar{X})(Y - \bar{Y})]$$

Coeficientul de covarianță dintre variabilele X și Y .

Prin $\bar{X}$ am notat $M(X)$ iar $\bar{Y} = M(Y)$.

Se poate arăta prin calcul că

$$\mu_{xy} = M(XY) - M(X) \cdot M(Y)$$

Coeficientul de corelație.

Numim **coeficient de corelație** a variabilelor aleatoare X și Y expresia:

$$\rho_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y} = \frac{M(XY) - M(X) \cdot M(Y)}{\sqrt{M(X^2) - M^2(X)} \cdot \sqrt{M(Y^2) - M^2(Y)}}$$

Proprietăți.

- 1) Coeficientul de corelație este un coeficient standardizat, cuprins între valorile -1 și $+1$.

$$-1 \leq \rho_{xy} \leq 1$$

- 2) Dacă variabilele X și Y sunt independente atunci

$$\rho_{xy} = 0$$

- 3) Dacă $\rho_{xy} = \pm 1$, între variabilele X și Y există o dependență liniară. Această dependență arată astfel:

$$y - M(Y) = \frac{\sigma_y}{\sigma_x} (x - M(X)), \text{ cazul } \rho = 1$$

$$y - M(Y) = -\frac{\sigma_y}{\sigma_x} (x - M(X)), \text{ cazul } \rho = -1$$

sau

$$x - M(X) = \frac{\sigma_x}{\sigma_y}(y - M(Y)), \text{ cazul } \rho = 1$$

$$x - M(X) = -\frac{\sigma_x}{\sigma_y}(y - M(Y)), \text{ cazul } \rho = -1$$

Observație. Dacă $\rho_{xy} = 0$ nu rezultă că variabilele X și Y sunt independente sau că există o slabă dependență între v.a. X și Y. Acest lucru poate să însemne mai degrabă că dependența nu este de tip liniar. În acest caz, se analizează cazurile de dependență parabolică, logaritmică, logliniară, etc.

Împrăștierea unei variabile aleatoare mai poate fi măsurată prin coeficientul de variație (sau de împrăștiere):

$$V = \frac{\sigma_x}{M(X)}$$

Aplicații.

1. Se atuncă 2 zaruri și se notează cu S numărul total de puncte care apar. Să se formeze tabloul distribuției lui S:

R.

$$S \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

2. Se dă variabila aleatoare X de distribuție:

$$X \begin{pmatrix} 0 & 1 \\ 0,3 & 0,7 \end{pmatrix}$$

Să se afle $X^2, X^3, \dots, X^n$.

$$\mathbf{R.} \quad X^n \begin{pmatrix} 0 & 1 \\ 0,3 & 0,7 \end{pmatrix}, \quad n \in \mathbb{N}$$

3. Fie v.a. X

$$X \begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Care este probabilitatea ca X să ia o valoare mai mică sau egală cu 3?

R.

$$P(X \leq 3) = P(X=1) + P(X=2) + P(X=3) = \frac{1}{2} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

4. Fie v.a. X, Y

$$X \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

$$Y \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Care este distribuția sumei $X+Y$?

R. A se vedea ex.1.

5. Fie X, Y 2 v.a. cu distribuțiile:

$$X \begin{pmatrix} -1 & 0 & 1 \\ p^2 & \frac{5}{3}p & \frac{1}{3} \end{pmatrix} \quad Y \begin{pmatrix} -1 & 0 & 1 & 2 \\ q^2 & \frac{8}{5}q & \frac{1}{6} & \frac{1}{30} \end{pmatrix}$$

Care este distribuția v.a $X+Y$ și XY ?

6. Fie

$$X \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0,3 & 0,4 & 0,2 & 0,1 \end{pmatrix}$$

Să se calculeze: $M(X)$, $M^2(X)$, $M(X^2)$, $M(X-1)$, $M(X^2-2X)$.

R.

$$M(X) = 0,3+0,8+0,6+0,4 = 2,1$$

$$M^2(X) = 4,41$$

$$M(X^2) = 1 \cdot 0,3 + 4 \cdot 0,4 + 9 \cdot 0,2 + 16 \cdot 0,1 = 5,3$$

$$X^2 \begin{pmatrix} 1 & 4 & 9 & 16 \\ 0,3 & 0,4 & 0,2 & 0,1 \end{pmatrix}$$

7. Fie

$$X \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{pmatrix}$$

Calculați dispersia lui X.

R.

$$m = M(X) = \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{2}{3} = 2,5$$

$$X-m \begin{pmatrix} -1,5 & -0,5 & 0,5 \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{pmatrix}$$

$$M[(X-m)^2] = 2,25 \cdot \frac{1}{6} + 0,25 \cdot \frac{1}{6} + 0,25 \cdot \frac{4}{6} = \frac{1}{6}(2,5+1) = \frac{35}{60}$$

8. Fie X o v.a. cu media m și dispersia σ^2 . Să se calculeze

valoarea medie și dispersia v.a. $Y = \frac{X-m}{\sigma}$.

R.

$$m(Y) = \frac{1}{\sigma} \sum_{i=1}^n (x_i - m) p_i = \frac{1}{\sigma} [\sum p_i x_i - m \sum p_i] = 0$$

$\sigma^2 = M[(X-m)^2]$. Să considerăm v.a. $(X-m)^2$ de distribuție:

$$X-m^2 \begin{pmatrix} (x_1-m)^2 & (x_2-m)^2 & \dots & (x_n-m)^2 \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Atunci

$$\begin{aligned}\sigma_y^2 &= \sum_{i=1}^n p_i (x_i - m)^2 = \sum_{i=1}^n p_i x_i^2 - 2m \sum_{i=1}^n p_i x_i + m^2 = \\ &= M(X^2) - 2m^2 + m^2 = M(X^2) - M^2(X) = \sigma_x^2\end{aligned}$$

9. O grupă de 58 de studenți susțin două examene la disciplinele “A” și “B”. Ne punem întrebarea dacă există vreo legătură între rezultatele obținute la cele două examene. Pentru aceasta, vom calcula coeficientul de corelație.

Distribuția rezultatelor studenților la cele două examene este prezentată în tabelul următor:

Note obținute la examenul “A”	Y	Note obținute la examenul “B”				
	X	10	9	8	7	Total
	10	2	2	0	0	4
	9	1	6	1	0	8
	8	0	4	15	4	23
	7	0	0	10	13	23
	Total	3	12	26	17	58

Construim variabilele X și Y atașate rezultatelor obținute la examenul “A” respectiv “B”.

$$X \begin{pmatrix} 10 & 9 & 8 & 7 \\ 4 & 8 & 23 & 23 \\ \hline 58 & 58 & 58 & 58 \end{pmatrix}$$

$$Y \begin{pmatrix} 10 & 9 & 8 & 7 \\ \frac{3}{58} & \frac{12}{58} & \frac{26}{58} & \frac{17}{58} \end{pmatrix}$$

$$m_1 = M(X) = \frac{4 \cdot 10 + 9 \cdot 8 + 8 \cdot 23 + 7 \cdot 23}{58} = \frac{457}{58} = 7,88$$

$$m_2 = M(Y) = \frac{10 \cdot 3 + 9 \cdot 12 + 8 \cdot 26 + 7 \cdot 17}{58} = \frac{465}{58} = 8,02$$

$$X^2 \begin{pmatrix} 10^2 & 9^2 & 8^2 & 7^2 \\ \frac{4}{58} & \frac{8}{58} & \frac{23}{58} & \frac{23}{58} \end{pmatrix}$$

$$Y^2 \begin{pmatrix} 10^2 & 9^2 & 8^2 & 7^2 \\ \frac{3}{58} & \frac{12}{58} & \frac{26}{58} & \frac{17}{58} \end{pmatrix}$$

$$M(X^2) = \frac{4 \cdot 10^2 + 8 \cdot 9^2 + 23 \cdot 8^2 + 23 \cdot 7^2}{58} = \frac{3647}{58} = 62,88$$

$$M(Y^2) = \frac{3 \cdot 10^2 + 12 \cdot 9^2 + 26 \cdot 8^2 + 17 \cdot 7^2}{58} = \frac{3769}{58} = 64,99$$

$$M(XY) = \frac{2 \cdot 100 + 2 \cdot 90 + 1 \cdot 90 + 6 \cdot 81 + 1 \cdot 72 + 4 \cdot 72 + 15 \cdot 64 + 4 \cdot 56 + 10 \cdot 56 + 13 \cdot 49}{58} = \frac{3697}{58} = 63,75$$

$$m_1^2 = M^2(X) = 7,88^2 = 62,09$$

$$m_2^2 = M^2(Y) = 8,02^2 = 64,32$$

$$\sigma_x^2 = 62,88 - 62,09 = 0,79; \quad \sigma_x = \sqrt{0,79} = 0,89$$

$$\sigma_y^2 = 64,99 - 64,32 = 0,67; \quad \sigma_y = 0,82$$

$$M(XY) - M(X) \cdot M(Y) = 63,75 - 63,20 = 0,55$$

Atunci

$$\rho_{xy} = \frac{M(XY) - M(X) \cdot M(Y)}{\sigma_x \cdot \sigma_y} = \frac{0,55}{0,73} = 0,75$$

Se constată astfel că între variabilele X și Y există o corelație directă destul de puternică.

Capitolul IV

INDICATORI AI CARACTERISTICILOR CANTITATIVE

Prin indicatori vom înțelege acele valori atașate variabilelor cantitative, care exprimă, sub formă sintetică, informația conținută în distribuția variabilei respective.

Există trei tipuri de indicatori:

- indicatori de poziție sau ai tendinței centrale de grupare
- indicatori de dispersie sau de împrăștiere
- indicatori ai formei distribuției

I. Indicatori ai tendinței centrale de grupare

Din seria indicatorilor de poziție sau al tendinței centrale de grupare, vom menționa:

1. Media aritmetică sau simplu media

Dacă $x_1, x_2, \dots, x_n$ sunt cele n valori pe care le poate lua o variabilă cantitativă, atunci valoarea mediei va fi:

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Exemplu. Să considerăm numărul familiilor dintr-un imobil după dimensiunea acestora (numărul de persoane ce alcătuiesc familia respectivă).

Nr. persoane	1	2	3	4	5	6	Total
Nr. familii	10	15	25	20	8	2	80

Se constată că numărul total de persoane este:

$$1 \cdot 10 + 2 \cdot 15 + 3 \cdot 25 + 4 \cdot 20 + 5 \cdot 8 + 6 \cdot 2 = 10 + 30 + 75 + 80 + 40 + 12 = 247$$

Atunci dimensiunea medie a familiei va fi $\frac{247}{80} = 3,087$

persoane / familie.

Desigur că nu poate exista o astfel de familie, dar acest indicator ne arată că în cazul familiilor absolut omogene, pe unde s-ar situa dimensiunea acestora..

În cazul unei variabile aleatoare discrete X

$$X \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0,05 & 0,07 & 0,08 & 0,13 & 0,32 & 0,22 & 0,13 \end{pmatrix}$$

valoarea medie este

$$\bar{X} = 1 \cdot 0,05 + 2 \cdot 0,07 + 3 \cdot 0,08 + \dots + 7 \cdot 0,13 = 4,78$$

iar în forma generală

$$X \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} \quad p_i \geq 0; \sum_{i=1}^n p_i = 1$$

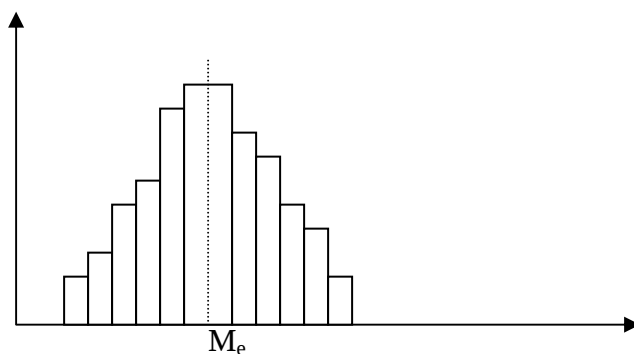
atunci $\bar{X} = M(X) = \sum_{i=1}^n p_i x_i$

Valoarea medie se mai numește și **speranța matematică** a variabilei X .

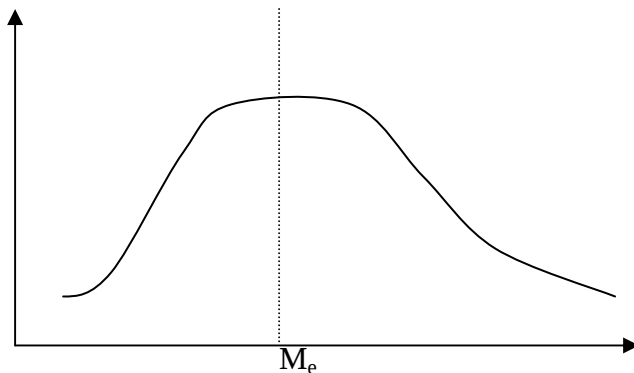
2. **Mediana** unei variabile cantitative X este acea valoare notată M_e a lui X pentru care are loc egalitatea:

$$P(X < M_e) = P(X > M_e) = \frac{1}{2}$$

Din punct de vedere grafic, mediana este acea valoare a lui X pentru care ariile din histogramă despărțite de ordonata lui M_e sunt egale.



Cazul variabilei discrete



Cazul variabilei continue

Dacă valoarea mediană M_e coincide cu o valoare x_i a variabilei X , atunci valoarea mediană este bine precizată. Dacă însă acest lucru nu se întâmplă, avem de-a face cu un **interval median**. În practică, se obișnuiește să se ia drept valoare a lui M_e mijlocul acestui interval.

Exemplu. Variabila atașată experienței de aruncare cu zarul are distribuția uniformă:

$$X \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

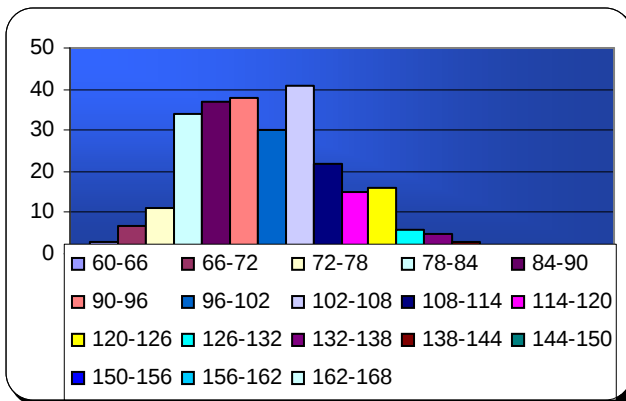
Intervalul median este $[3,4]$ iar valoarea medianei M_e va fi:

$$M_e = \frac{3+4}{2} = 3,5$$

3. **Modul** sau valoarea dominantă este în același timp și valoarea cea mai probabilă pe care o poate lua variabila X . Se notează prin M_0 sau $\hat{X}$.

Exemplu. Fie o variabilă cantitativă continuă, de exemplu dimensiunile unor piese, care au fost observate ca variind între 60mm și 168mm. Acest interval a fost împărțit din motive practice, în intervale de 6mm, obținându-se discretizarea variabilei și următoarea tabelă:

Intervale	Centrul intervalelor	Frecvența de aparitie
60-66	63	3
66-72	69	7
72-78	75	11
78-84	81	34
84-90	87	37
90-96	93	38
96-102	99	30
102-108	105	41
108-114	111	22
114-120	117	15
120-126	123	16
126-132	129	6
132-138	135	5
138-144	141	3
144-150	147	1
150-156	153	0
156-162	159	0
162-168	165	1
		Total 270



Frecvența maximă se obține pentru $x=105$, dar din examinarea histogramei, se constată că această frecvență maximă pare a fi accidentală în examinarea tendinței generale a fenomenului statistic, și ca atare, ar putea fi datorată faptului că în eșantionarea a 20 de observații studiate, hazardul a grupat în intervalul 102-108 o fracțiune mai importantă decât aceea care se găsește în mod normal în populația statistică.

Trasarea curbei frecvențelor implică, pe lângă continuitatea fenomenului și o formă potrivită, clasică, astfel încât aria totală să fie aceeași, prin compensare.

În acest exemplu, histograma indică de fapt, ca **modul**, valoarea 95. Curba se efectuează printr-o **ajustare analitică**.

Între valorile:

M_0 - modul sau valoare modală,

M_e - valoarea mediană

$\bar{x}$ - media

există o relație aproximativă, valabilă pentru distribuții cel mult ușor asimetrice:

$$M_0 = 4M_e - 3\bar{x}$$

Valorile celor trei indicatori ai tendinței centrale sunt folosite pentru construirea parametrilor care redau forma distribuției.

II Indicatori de dispersie

Indicatorii de dispersie caracterizează o populație statistică din punctul de vedere al omogenității (eterogenității), în raport cu o variabilă cantitativă dată.

În anumite situații, indicatorii de dispersie pot reflecta **gradul de inegalitate** între indivizii statistici, în raport cu o anumită caracteristică.

În **modelele explicative**, indicatorii de dispersie pot explica **gradul de nedeterminare**, de variabilitate al unui fenomen.

Amplitudinea

Este diferența dintre cea mai mare și cea mai mică valoare.

$$A = x_{\max} - x_{\min}$$

Quantile

Fie X o variabilă aleatoare al cărui argument x este definit în intervalul $[a, b]$. S-a văzut că pentru determinarea medianei M_e trebuie rezolvată ecuația

$$F(x) = \frac{1}{2}$$

unde $F(x) = P(X < x)$ este funcția de repartiție a variabilei X .

Numim **quantile** de ordinul n ale variabilei X , rădăcinile ecuației:

$$F(x) = \frac{i}{n}, \quad i = 1, 2, \dots, n-1$$

pentru $n \in \mathbb{N}$ dat, iar $F(x)$ este funcția de repartiție.

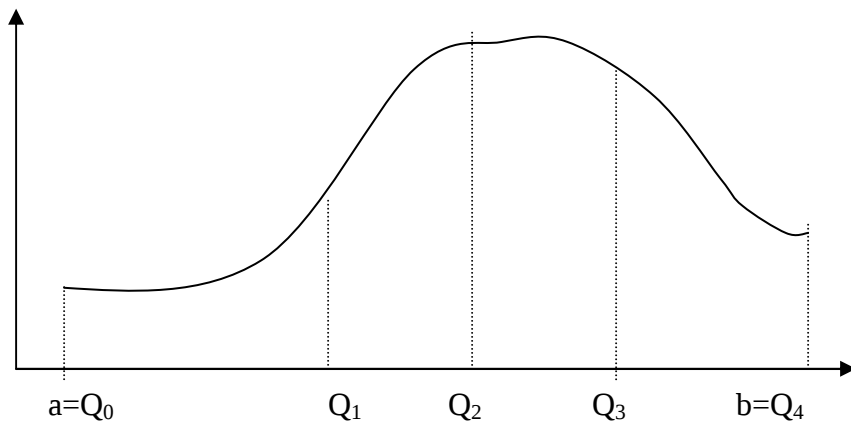
Pentru $n = 2$ se obține mediana M_e .

Pentru $n = 4$, cele 3 rădăcini: Q_1, Q_2, Q_3 se vor numi **cuartile**

Pentru $n = 10$ soluțiile se numesc **decile**

Pentru $n = 100$ soluțiile se numesc **centile**

Dacă reprezentăm grafic curba de distribuție, quantilele de ordinul n împart suprafața mărginită de curba de distribuție, axa OX și ordonatele $x = a, x = b$ în n părți de arii egale (sau împart mulțimea indivizilor în n părți egale).



În cazul cuartilelor ($n = 4$), a doua cuartilă este egală cu mediana:

$$Q_2 = M_e$$

- Q_1 se mai numește cuartila mică sau inferioară;
- Q_3 cuartila mare sau superioară.

Diferența:

$$I = Q_3 - Q_1$$

se numește **abaterea intercuartilă** (sau abaterea cuartilă).

Valoarea:

$$\frac{Q_3 - Q_1}{2}$$

se va numi **abatera semiintercuartilă**

Adesea se folosește o valoare relativă (standardizată)

$$\frac{Q_3 - Q_1}{Q_2}$$

numită **abatera intercuartilă relativă**.

Să considerăm rezultatele obținute pe un lot de 1000 persoane la un test cu valori de 1 la 10.

Valori	1	2	3	4	5	6	7	8	9	10	Total
Frecvențe simple	15	25	90	120	200	220	160	110	40	20	1000
Frecvențe cumulate	15	40	130	250	450	670	830	940	980	1000	

Amplitudinea: $10 - 1 = 9$

Prima cuartilă, se obține prin delimitarea primilor 250 de indivizi este 4, **mediana** este 6 (deoarece sub 5 sunt 450 de indivizi iar sub 6 sunt 670):

A **treia cuartilă** este 7 (sub 7 sunt 830 de indivizi, deci și al 750 -lea). Așadar,

$$Q_1 = 4$$

$$Q_2 = M_e = 6$$

$$Q_3 = 7$$

$$I = Q_3 - Q_1 = 7 - 4 = 3$$

$$I_{\text{rel}} = \frac{Q_3 - Q_1}{Q_2} = \frac{7 - 4}{6} = 0,5$$

Utilizarea decilelor și a centilelor se practică pentru a măsura inegalitățile dintre oameni.

Exemplu. În cazul venitului (pe familie sau pe cap de locuitor, etc.) se calculează venitul mediu al primilor 10% (cei mai bogați) și venitul mediu al ultimilor 10% (cei mai săraci) și se compară cele două valori, printr-un indice standardizat.

Indicele lui Gini

Carrado Gini a propus un indice ca fiind media aritmetică a diferențelor dintre toate perechile de valori luate în valoare absolută (fără sume):

$$G = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|$$

pentru $i \neq j$, în cazul caracteristicilor fără frecvență, sau

$$G = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| f_i f_j$$

pentru $i \neq j$, în cazul caracteristicilor cu frecvență (în care f_i este frecvența relativă a valorii x_i)

Exemplu. Într-o sesiune, un student a obținut la 5 examene notele 6,7,8,9,10. Pentru a calcula indicele lui Gini, vom lua în calcul diferențele în modul:

$$|6-7|, |6-8|, |6-9|, |6-10|, |7-6|, |7-8|, |7-9|, |7-10|, \\ |8-6|, |8-7|, |8-9|, |8-10|, |9-6|, |9-7|, |9-8|, |9-10|, \\ |10-6|, |10-7|, |10-8|, |10-9|$$

$$S = 1 + 2 + 3 + 4 + 1 + 1 + 2 + 3 + 2 + 1 + 1 + 2 + 3 + 2 + 1 + 1 + 4 + \\ + 3 + 2 + 1 = 40$$

$$\text{Deci: } G = \frac{1}{4 \cdot 5} \cdot 40 = 2$$

Această valoare ne spune că diferența medie între două valori diferite este de 2.

Abaterea medie.

Dacă a este o constantă, atunci mărimea

$$A_M(a) = \frac{1}{n} \sum_{i=1}^n |x_i - a| \text{ pentru o serie de valori individuale}$$

sau

$$\frac{\sum_{i=1}^n f_i |x_i - a|}{\sum_{i=1}^n f_i} \text{ pentru o repartiție de frecvențe se va numi}$$

abaterea medie de la a . Dacă $a = m = M(x)$, atunci $A_M(m)$ este abaterea medie de la media lui X , sau mai simplu, **abaterea medie.**

Yule și Kendall au arătat că cea mai mică abatere medie se obține atunci când se ia drept constantă a valoarea medianei Me. În exemplul anterior, Me = 8 și abaterile de la mediană vor fi:

$$|6-8|=2; \quad |7-8|=1; \quad |8-8|=0; \quad |9-8|=1; \quad |10-8|=2;.$$

Media acestor 5 valori va fi:

$$\frac{2+1+0+1+2}{5} = \frac{6}{5} = 1,2$$

Media celor 5 note va fi:

$(6+7+8+9+10):5=8$, iar abaterea de la medie:

$(2+1+0+1+2):5=1,2$. Această valoare coincide cu valoarea obținută cu calculul medianei deoarece în acest caz mediana și valoarea medie au aceeași valoare.

Abaterea pătratică medie (abatere standard, abaterea tip, σ).

Abaterea pătratică medie este rădăcina pătrată din media aritmetică a pătratelor abaterilor valorilor observate în raport cu media lor aritmetică:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \text{ pentru o serie de valori individuale și}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}} \text{ pentru o repartiție de frecvențe.}$$

Expresia

$$\sigma^2 = M[(x - m)^2] = M[(x - \bar{x})^2]$$

se mai numește **dispersie** sau **varianță**.

Dacă populația statistică este concepută ca o mulțime de grupuri, atunci media generală a caracteristicii este egală cu media mediilor fiecărui grup.

Varianța σ_x^2 se va numi în acest caz **varianță intergrupală**:

$$\sigma_x^2 = \frac{1}{n} \sum_{j=1}^s n_j (\bar{x}_j - \bar{x})^2$$

unde s este numărul grupurilor, iar

$$n_1 + n_2 + \dots + n_s = n$$

Se poate calcula o medie a varianțelor din cadrul grupului, notată $\bar{\sigma}$, numită **varianță intragrupală** dată de formula:

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n n_j \sigma_j^2$$

Se demonstrează că:

$$\sigma^2 = \bar{\sigma}^2 + \sigma_x^2$$

adică “varianța totală” se descompune în suma dintre varianța intragrupală și a celei intergrupale.

Coeficient de variație (al lui Pearson)

Raportul dintre abaterea standard și media variabilei X se va numi **coeficient de variație**:

$$v_x = \frac{\sigma_x}{\bar{x}}$$

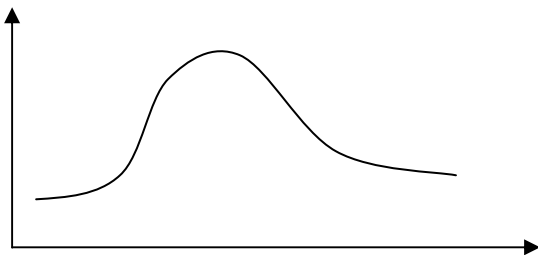
III. Indicatori ai formei distribuției.

Forma distribuției unei caracteristici cantitative este măsurată de doi indicatori:

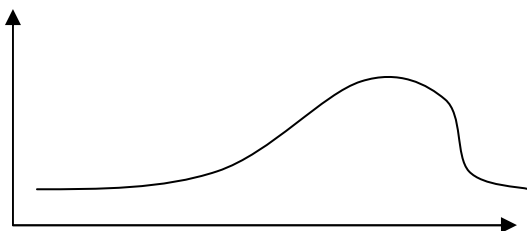
$$1. \text{ Oblicitatea} = \frac{\bar{x} - M_0}{\sigma} = \frac{3(\bar{x} - Me)}{\sigma}$$

(Formula lui Pearson)

Dacă această expresie are semn pozitiv curbele sunt alungite către dreapta:



iar când expresia este negativă, alungirea este spre stânga:

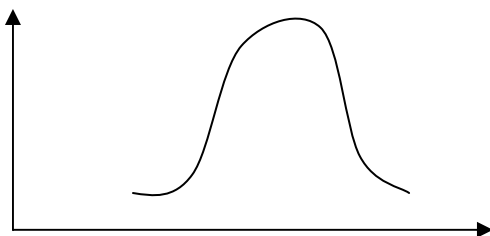


2. Indicatori de boltire.

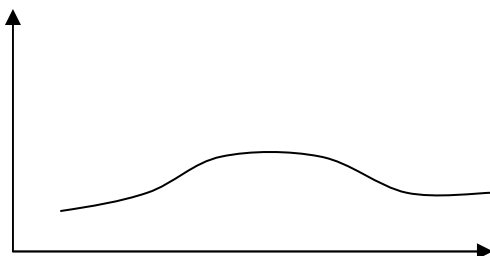
Acest indicator are expresia:

$$B = \frac{1}{n\sigma^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3$$

și este pozitiv în cazul boltirilor pronunțate:



și negativ în cazul boltirilor aplatizate:



Calculul acestor indicatori se poate executa cu ajutorul calculatorului electronic pe baza programului “SPSS”.

Capitolul V

CORELAȚIA RANGURILOR.

Să presupunem că avem o serie de n unități statistice:

$$U_1, U_2, \dots, U_n$$

fiecare dintre acestea având două caracteristici

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

De exemplu, aceste n unități statistice pot reprezenta n persoane iar caracteristicile pot fi înălțimea și greutatea celor n persoane.

Problema care se pune este dacă există o corelație între aceste două caracteristici.

Să admitem că s-a făcut următoarea înregistrare a datelor pe un lot de 10 persoane și că această înregistrare s-a făcut după două caracteristici:

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
(1)	3	1	10	6	2	5	7	8	4	9
	1	2	9	7	4	3	6	10	5	8

Astfel, în înregistrarea (1) am așezat în prima linie cele 10 persoane.

În linia a doua am înregistrat locul pe care îl ocupă fiecare persoană în raport cu prima caracteristică, iar pe linia a treia,

aceeași ordine referitoare la a doua caracteristică. De exemplu, persoana U_1 este al treilea în ordinea crescătoare a primei caracteristici (care poate fi înălțimea) și are locul întâi în ordinea crescătoare a celei de a doua caracteristici (de ex. greutatea).

Dacă am fi avut ordinea (2):

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
(2)	3	1	10	6	2	5	7	8	4	9
	3	1	10	6	2	5	7	8	4	9

înseamnă că am fi avut cea mai strânsă legătură între cele două caracteristici. În acest caz, fiecare persoană U_i ocupă același loc în ordinea ierarhică a celor două caracteristici.

Dacă, dimpotrivă, am fi avut ordinea (3)

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
(3)	3	1	10	6	2	5	7	8	4	9
	8	10	1	5	9	6	4	3	7	2

atunci am fi avut **discordanță** maximă între aceste două caracteristici.

Dacă considerăm clasificarea (1), diferențele dintre linia II și linia III sunt:

$$d_i: 2, -1, 1, -1, -2, 2, 1, -2, -1, 1$$

Se observă că : $d_1 + d_2 + \dots + d_{10} = 0$.

De altfel, în general

$$\sum_{i=1}^n d_i = 0$$

Cu cât diferențele d_i sunt mai mari în valoare absolută, cu atât avem o discordanță mai mare între caracteristici.

Astfel, este propus următorul coeficient:

$$\rho = 1 - 6 \frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (n > 1)$$

numit coeficient de **corelație a rangurilor** al lui **Spearman**. În cazul exemplului de mai sus,

$$\sum_{i=1}^{10} d_i^2 = 22.$$

Prin urmare:

$$\rho = 1 - \frac{6 \cdot 22}{10^3 - 10} = 0,867$$

Se știe că $\rho \in [-1, 1]$, ceea ce înseamnă că acest coeficient ia valori cuprinse între -1 și 1 . Când $\rho = 1$ avem concordanță maximă între cele două variabile, iar când $\rho = -1$ avem discordanță maximă între cele două variabile.

Valoarea $\rho = 0,867$ fiind destul de apropiată de 1 , arată că între cele două caracteristici există o corelație destul de strânsă.

Coeficientul de corelație a rangurilor al lui Kendall

Să reluăm exemplul anterior și să luăm în considerație următoarea clasificare:

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
(4)	1	2	3	4	5	6	7	8	9	10
	2	4	1	5	3	7	6	10	8	9

Tabloul (4) conține aceleași elemente ca și clasificarea (1), cu diferența că în linia a doua rangurile au fost scrise în ordine crescătoare de la 1 la 10, iar persoanele U_i ($i = 1, \dots, 10$) au aceleași caracteristici în ambele clasificări.

Să constatăm că prima persoană din înregistrarea (4) ocupă locul 1 în ce privește prima caracteristică și locul 2 în ce privește a doua caracteristică. Există deci 9 persoane care depășesc pe U_2 în ce privește prima caracteristică și numai 8 care îl depășesc pe U_2 în raport cu a doua caracteristică.

Practic, obținem următorii indicatori

8, 6, 7, 5, 5, 3, 3, 0, 1

astfel:

locul 2 de pe linia a 3-a din tabloul (4) este depășit de 8 poziții. Locul 4 de 6 poziții, locul 1 de 7 poziții, ș.a.m.d. Vom nota cu P suma numerelor mai mari (care depășesc o anumită poziție) și prin Q suma numerelor mai mici (care sunt depășite de acea poziție).

Astfel:

$$P = 8 + 6 + 7 + 5 + 5 + 3 + 3 + 0 + 1 = 38$$

$$S = 1 + 2 + 0 + 1 + 0 + 1 + 0 + 2 + 0 = 7$$

Kendall a propus următorul indicator pentru concordanța rangurilor:

$$r = \frac{2(P - Q)}{n(n - 1)}$$

Când $r = 1$ obținem concordanță maximă iar când $r = -1$ se obține discordanță maximă. Când $r = 0$ are loc independența între cele două caracteristici.

În cazul exemplului nostru

$$r = \frac{2(38 - 7)}{10 \cdot 9} = 0,69$$

ceea ce indică o concordanță pozitivă.

P se mai numește **indicator al concordanței pozitive** deoarece el crește odată cu creșterea lui r în timp ce Q se va numi **indicator al concordanței negative**, întru-cât r descrește când el crește.

Acest coeficient al lui Kendall se aplică numai pentru serii mari. Pentru serii mici, dispersia acestui coeficient are valori mari.

Capitolul VI

ANALIZA DE REGRESIE

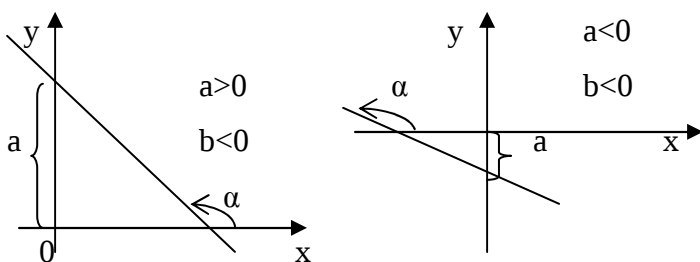
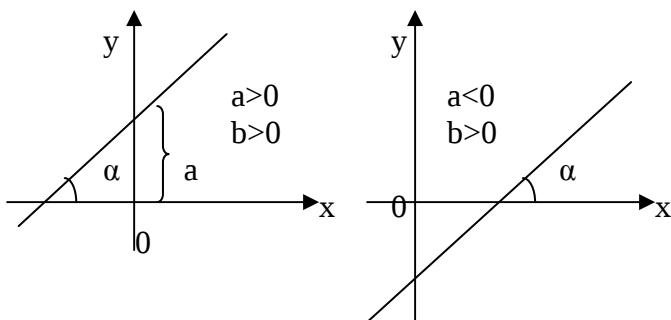
În statistica socială întâlnim adesea repartiții în care fiecărei unități a populației îi corespund simultan două sau mai multe caracteristici. Astfel de repartiții se mai numesc bidimensionale sau multidimensionale. Ele ne pot sugera existența unor relații între caracteristicile respective.

Prezența sau absența unor astfel de relații, ca și amploarea acestora, formează obiectul analizei seriilor interdependente. Ea presupune analiza simultană a două variabile și folosește două tipuri de metode statistice: **regresia** și **corelația**.

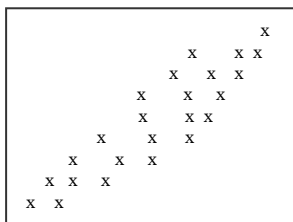
De regulă, una dintre aceste două variabile este considerată ca variabilă independentă sau explicativă, în timp ce a doua este o variabilă dependentă. Acest lucru are loc dacă a doua variabilă prezintă modificări la variațiile primei variabile. Analiza acestei dependențe se face prin metoda **regresiei**. Termenul de regresie a fost dat de statisticianul englez Francis Galton (1822-1911) și el stabilește legătura care există între cele două variabile X și Y. Cazul cel mai simplu de regresie este cel liniar. Asta înseamnă că punctele $A_i(x_i, y_i)$ se distribuie în jurul unei drepte:

$$y = a + bx \tag{1}$$

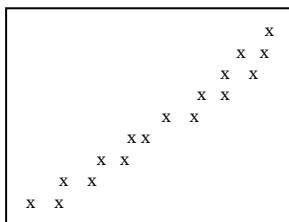
Reprezentate în plan într-un sistem de axe coordonate, o astfel de legătură de tip liniar poate avea una din următoarele forme:



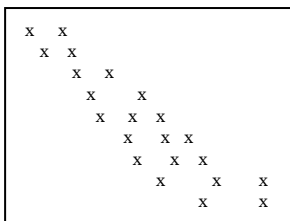
Acestor tipuri de legături liniare le corespund diagramele de împrăștiere a valorilor celor două variabile:



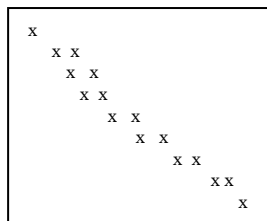
1)



2)



3)



4)

Determinarea parametrilor a și b din ecuația dreptei de regresie se face cu ajutorul **metodei celor mai mici pătrate** care spune că “suma pătratelor diferențelor dintre valorile empirice $\tilde{y}_i$ și valorile teoretice y_i date de ecuația de regresie, să fie minimă. Grafic, pătratul diferențelor dintre $\tilde{y}_i$ și y_i se exprimă prin aria unor pătrate ale căror laturi sunt egale cu $\tilde{y}_i - y_i$. Suma ariilor acestor pătrate va fi cu atât mai mică cu cât valorile empirice ale lui y se vor apropia mai mult de valorile teoretice corespunzătoare

$$S = \sum_{i=1}^n (\tilde{y}_i - y_i)^2 = \text{minim} \quad (2)$$

și ținând cont de formula (1), se obține condiția:

$$\sum_{i=1}^n (\tilde{y}_i - a - bx_i)^2 = \text{minim} \quad (3)$$

Utilizând metode ale analizei matematice, anularea derivatelor parțiale în (3), se obține sistemul în necunoscutele a și b :

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases} \quad (4)$$

a cărui rezolvare conduce la soluțiile:

$$a = \frac{m_2 \sigma_1 - m_1 \rho \sigma_2}{\sigma_1} \text{ și } b = \frac{\rho \sigma_2}{\sigma_1} \quad (5)$$

Se obține astfel dreapta de regresie

$$y - m_2 = \frac{\rho \sigma_2}{\sigma_1} (x - m_1) \quad (6)$$

unde m_1 , m_2 sunt $M(X)$, $M(Y)$ – adică mediile variabilelor X și Y , σ_1 , σ_2 - dispersiile lor iar ρ - coeficientul de corelație.

În mod analog, dacă ne propunem să determinăm parametrii arbitrari a și b astfel încât

$$M(x - a - by)^2 = \sum_{i=1}^n (x_i - a - by_i)^2 = \text{minim}$$

vom putea obține o altă dreaptă de regresie:

$$y - m_2 = \frac{\rho \sigma_1}{\sigma_2} (x - m_1) \quad (7)$$

Cele două drepte de regresie date de relațiile (6) și (7) se intersectează în punctul de coordonate $G(m_1, m_2)$ care se va numi **centrul de greutate** al distribuției.

În general, aceste drepte de regresie sunt diferite, afară de cazul când

$$\frac{\rho\sigma_1}{\sigma_2} = \frac{\rho\sigma_2}{\sigma_1}$$

sau $\sigma_1 = \sigma_2$

Prin urmare, cele două drepte de regresie coincid dacă dispersiile σ_1 și σ_2 sunt egale.

Să luăm drept exemplu, două variabile X și Y pentru care au fost înregistrate 15 observații conform cu tabelul de mai jos:

	X_i	Y_i	X_i^2	X_iY_i
1	1	1	1	1
2	1	-5	1	-5
3	1	-9	1	-9
4	3	8	9	24
5	3	1	9	3
6	3	-3	9	-9
7	4	11	16	44
8	4	3	16	12
9	4	0	16	0
10	14	17	196	238
11	14	12	196	168
12	14	9	196	126
13	33	26	1089	858
14	33	19	1089	627
15	33	17	1089	561
Σ	165	107	3933	2639

Tabelul 1

Ecuția de regresie are forma:

$$Y = a + bX$$

unde
$$a = \frac{\sum x_i \sum x_i y_i - \sum x_i^2 \sum y_i}{(\sum x_i)^2 - n \sum x_i^2}$$

și

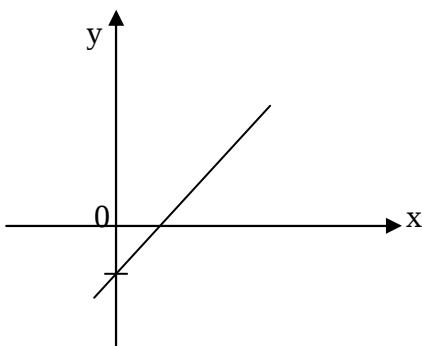
$$b = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2}$$

Conform datelor din tabelul 1 obținem:

$a = -0,45$; $b = 0,69$, deci

$$Y = -0,45 + 0,69X$$

cu reprezentarea grafică:



Capitolul VII

ANALIZA DE DEPENDENTA

Conceptul de **analiză de dependență** sau cum era cunoscut anterior, **path analysis** sau **cauzal analysis** se referă la determinarea relațiilor între un ansamblu de variabile în contextul unei structuri cauzale, adică o structură a unui grup de variabile între care se constată sau se presupune anumite relații de interdependență.

Prin **variabilă** vom înțelege orice criteriu de clasificare, fie că este vorba de o însușire dihotomică (sexul), de ordine (nivelul de școlarizare) sau cantitativă (venitul).

Principalele modele propuse de Simon, Blalock și R.Boudon se bazează pe observația coeficientului de corelație liniară între variabile și nu se depărtează de tehnicile clasice ale analizei de regresie.

Vom nota prin X un ansamblu de variabile

$$X = (x_1, x_2, \dots)$$

Un model de structură cauzală va fi o structură cauzală în care ipotezele sunt făcute pe baza notării relațiilor între variabile.

Modelul recursiv.

Acest model a fost studiat și dezvoltat de M.Simon, H.Blalock și R. Boudon. De el s-au mai ocupat Duncan și Alker.

Ipotezele modelului recursiv.

Ipoteza 1. Relațiile dintre variabile sunt liniare

Asta înseamnă că orice variabilă a grafului este exprimată ca funcție liniară de una sau mai multe variabile care o precede în graf.

În acest caz, graful constituie reprezentarea grafică a relațiilor analitice care definește structura cauzală.

De exemplu, sa presupunem ca avem un model sub formă de graf orientat cu patru variabile x_1 , x_2 , x_3 , x_4 , în care săgețile arată influențele exercitate de unele variabile asupra altora.

În fig. alăturată avem un astfel de model:

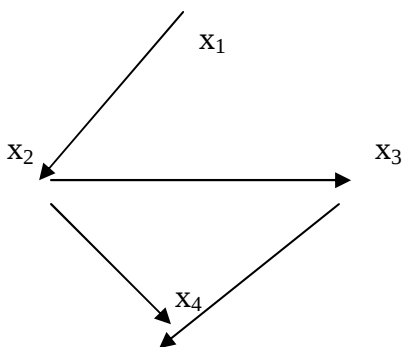


Fig. 1

Faptul că variabila x_2 este determinată de variabila x_1 , îl vom scrie analitic astfel:

$$x_2 = a_{12}x_1 + e_2 \quad (1)$$

adică x_2 este funcție liniară de o singură variabilă x_1 ; e_2 – măsoară reziduul, adică abaterea dintre valoarea variabilei x_2 și cantitatea explicată de x_2 ; M.Simon îl numește termen de eroare iar R. Boudon îl numește factor care acționează implicit asupra lui x_2 . Graful asociat relației (1) este în acest caz:

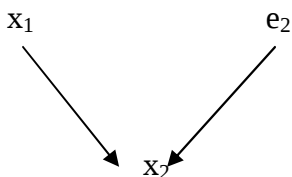


Fig.2

Dacă se aplică acestui model metoda celor mai mici pătrate a lui Gauss, atunci coeficientul a_{12} poate fi privit drept coeficientul de regresie al lui x_2 în raport cu x_1 .

A doua ecuație din modelul analitic asociat grafului din fig.1, va fi:

$$x_3 = a_{23}x_2 + e_3, \quad (2)$$

dacă variabila x_3 este funcție liniară numai de x_2 .

În fine, x_4 este funcție de x_2 și de x_3 deci:

$$x_4 = a_{24}x_2 + a_{34}x_3 + e_4 \quad (3)$$

Nu există termen în x_1 pentru că nu există săgeată între x_1 și x_4 .

Să mai observăm că în modelul din fig.1, x_1 este o variabilă primară, ea nefiind influențată de o altă variabilă din sistem, în timp ce variabilele x_2 , x_3 și x_4 sunt variabile dependente.

Ipoteza 2. Nu există efect de interacțiune. Acest lucru înseamnă că relația dintre două variabile nu este funcție de o a treia.

Un exemplu sugestiv este dat de Boudon și Lazarsfeld, care studiază influența vârstei și a nivelului educației școlare asupra receptării radiofonice a muzicii clasice.

Luată separat, fiecare variabilă nu prezintă nici o relație particulară cu ascultarea muzicii clasice. Pe de altă parte, dacă s-ar studia subpopulația adultă, s-ar constata că există o relație între receptarea muzicii clasice și nivelul de educație. Cei cu un nivel de instruire mai ridicat receptează mai mult emisiunile de muzică clasică decât alte tipuri de emisiuni.

Această a doua ipoteză este cumva cuprinsă în prima.

Ipoteza 3. Reziduurile e_i nu sunt corelate între ele.

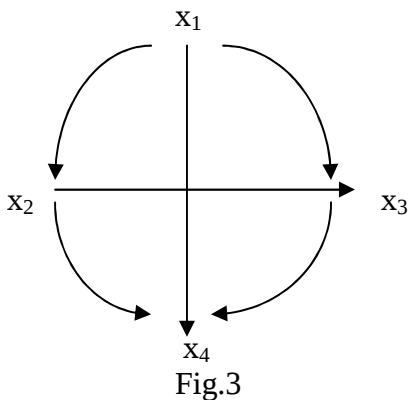
Asta înseamnă că în graf nu există săgeți între e_i pe de o parte, iar pe de altă parte nu există săgeți nici între e_i și x_j pentru $i \neq j$.

Construcția unui model de structură cauzală.

Să considerăm grupul de 4 variabile ordonate:

x_1, x_2, x_3, x_4

și să constuim un graf complet, adică orice cuplu de variabile este legat print-o săgeată:



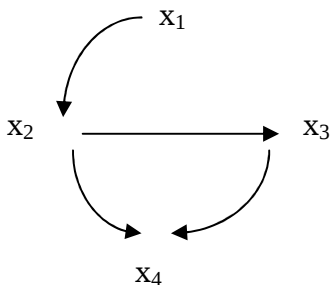
iar sistemul asociat grafului din fig.3 va fi:

$$x_2 = a_{12}x_1 + e_2$$

$$x_3 = a_{13}x_1 + a_{23}x_2 + e_3$$

$$x_4 = a_{14}x_1 + a_{24}x_2 + a_{34}x_3 + e_4$$

Făcând ipoteze simplificatoare privind prezența sau absența relațiilor între variabile, vom putea construi un graf derivat din primul, obținut deci prin suprimarea anumitor săgeți.



Atașăm grafului derivat din fig.4 sistemul de ecuații corespunzător:

$$x_2 = a_{12}x_1 + e_2$$

$$(*) \quad x_3 = a_{23}x_2 + e_3$$

$$x_4 = a_{24}x_2 + a_{34}x_3 + e_4$$

După cum se poate constata, absența unei săgeți în graful orientat este echivalentă cu anularea coeficientului de regresie a_{ij} corespunzător.

Analiza și testarea modelelor matematice a structurilor cauzale se poate face în multe feluri.

O cale este aceea prin care verificăm dacă coeficienții a_{ij} corespunzători săgeților absente, sunt nuli, metodă utilizată de H.Blalock. Altfel, putem rezolva sistemul (*), pentru a-l urma pe R.Boudon.

Modelul lui Blalock

După construirea structurii cauzale și a sistemului de ecuații asociat, Blalock ia în considerație numai coeficienții a_{ij} care sunt nuli, ca urmare a ipotezelor inițiale.

Acești coeficienți sunt coeficienții de regresie parțială din ecuația analizei de regresie care este avută în considerație.

Astfel, în schema din fig.4, care are asociat sistemul (*), avem:

$$a_{13} = b_{13,2}$$

care este coeficientul de regresie parțială între x_1 și x_3 ; deci $a_{13} = 0$ este echivalent cu $b_{13,2} = 0$. Pe de altă parte

$$b_{13,2} = r_{13,2} \cdot \frac{s_{1,2}}{s_{3,2}}$$

care leagă coeficientul de regresie parțială de cel de corelație parțială. De aici rezultă $r_{13,2} = 0$.

În continuare, se poate proceda în două moduri:

1. Se calculează acest coeficient pe baza datelor empirice, pentru a se vedea dacă el este apropiat de zero (în practică este puțin probabil ca el să fie egal exact cu zero!)
2. Să se deducă o relație între coeficienții de corelație. Astfel, $r_{13,2} = 0$ antrenează $r_{13} = r_{12} \cdot r_{23}$ deci produsul $r_{12} \cdot r_{23}$ calculat dă o valoare teoretică a lui r_{13} , pe care o notăm $\tilde{r}_{13}$; această valoare poate fi comparată cu valoarea obținută empiric a lui r_{13} ; calculată direct pe baza datelor.

În cazul modelului lui Blalock, în situația când anumite legături cauzale sunt presupuse nule, problema care se pune este de a ști ce se întâmplă dacă una dintr-ele este greșit pusă.

Astfel, dacă ipoteza care conduce la o anumită ecuație de ex. $r_{13,2} = 0$, este greșită, ar trebui să ne așteptăm ca ansamblul valorilor teoretice care au intrat în această relație, să se îndepărteze sensibil de valorile empirice corespunzătoare.

Dacă însă ipoteza pusă legată de o verigă intermediară este falsă, numai acest din urmă coeficient va prezenta o variație sensibilă în raport cu valoarea empirică corespunzătoare, fără ca ecuațiile și deci legăturile anterioare să fie afectate.

În ambele cazuri, eroarea este de aceeași natură. Dacă o relație este falsă, adăugarea săgeților corespunzătoare poate să îmbunătățească modelul propus.

În situația în care o greșeală se repercutează asupra mai multor relații, vom avea tendința să respingem modelul. Dacă eroarea rămâne localizată asupra unei singure relații, modelul poate fi recuperat, indicându-se astfel și modalitățile de ameliorare.

Din punct de vedere metodologic, un alt punct controversat al acestui model constă în aceea că nu este propus nici un test care să indice gradul de apropiere a valorilor teoretice de cele empirice. În acest sens, cercetătorul are mai multă libertate de decizie, de validare sau invalidare a modelului propus, de îmbunătățire a acestuia atunci când rezultatele obținute nu au fost satisfăcătoare.

Construcția apriorică a structurilor cauzale recursive deduse din ipoteze sau rezultate din teorie, ar trebui să fie urmată de o analiză profundă a validității prezenței sau absenței fiecăreia dintre relațiile date. Această analiză ar consta dintr-un șir de analize de regresie, aplicabile primelor variabile ale modelului, apoi variabilelor intermediare introduse pe parcursul analizei. Metodologia aceasta oferă posibilitatea și chiar oportunitatea revenirii periodice asupra ipotezelor modelului și eventuala ameliorare a acestora.

Totodată este lăsată la latitudinea și competența cercetătorului decizia de a se introduce o nouă legătură cauzală în modelul explicativ propus și în ce măsură această decizie are rolul de a îmbunătăți modelul propus.

Capitolul VIII

SONDAJUL STATISTIC ȘI EȘANTIONUL STATISTIC

Ce este sondajul statistic?

Procedeul statistic prin care reușim să prelevăm o parte din populația intrată în studiu, denumit **univers de eșantionare**, se va numi **sondaj statistic** sau **selecție**.

Rezultatul acestei operațiuni se va numi eșantion.

Așadar, eșantionul este subcolectivitatea extrasă din populația totală, pe care dorim să o studiem. Neavând intenția (și adesea nici posibilitatea) de a studia fiecare unitate statistică a întregii populații, scopul nostru este să alegem eșantionul de așa manieră, încât studiindu-l, rezultatele obținute să poată fi extinse asupra întregii populații din universul de eșantionare.

În preajma alegerilor electorale locale sau generale, candidații și formațiunile politice interesate, doresc să afle cum se poziționează ierarhic în preferințele electoratului. În cazul alegerilor generale, universul de eșantionare înseamnă populația adultă a țării, adică de peste 18 ani împliniți. Asta înseamnă aproximativ 16 milioane de alegători. Statistica reușește ca, prin respectarea riguroasă a anumitor reguli, să

extragă un eșantion de 1200-1800 de subiecți, iar rezultatele obținute din anchetarea acestor subiecți, să coincidă, în limita unei erori minime, controlate, cu rezultatele care s-ar fi obținut dacă ar fi fost anchetată întreaga populație adultă. Acest lucru este cu adevărat remarcabil.

Deși cu vechi tradiții, sondajul statistic a căpătat notorietate științifică în prima jumătate a secolului 20. El a fost aplicat cu mult succes în perioada interbelică prin anchetele și rezultatele obținute de George Gallup în SUA și mai apoi în Franța și Anglia anilor premergători celui de-al doilea război mondial.

În aceeași perioadă s-au realizat progrese notabile în teoria și practica sondajelor statistice, prin contribuțiile aduse de lucrările lui J.Neyman privind eșantionarea simplă aleatoare, stratificată, multistadială, construirea optimală a unui eșantion stratificat, etc.

Odată puse bazele teoriei sondajelor, cercetările s-au îndreptat în direcția problemelor practice care le ridică utilizarea sondajelor: metode de eșantionare, de estimare, de alcătuire a chestionarelor de opinie, de instruire a operatorilor de teren, codificarea și prelucrarea datelor recoltate, interpretarea rezultatelor obținute.

Toate acestea au fost posibile ca urmare a dezvoltării și utilizării metodelor statistice, cu precădere a statisticii inferențiale.

Contribuții esențiale în acest domeniu și-au adus K.Pearson, J.Neyman, A.N.Kolmogorov, R.Fisher, Feller, Gnedenko, etc.

În anii din urmă, deși practica sondajelor statistice a luat o amploare fără precedent, utilizarea metodelor statistice a fost utilizată cu stângăcie sau în mod neadecvat. Acest lucru se poate adesea întâlni și în sondajele de opinie care au caracter preelectoral și care confruntate la scurt timp cu realitatea – care a constituit-o alegerile locale sau generale, pune în evidență uneori mari abateri de la realitate.

Aceste situații neplăcute pot fi evitate prin cercetări minuțioase, prin amplasarea judicioasă în teren a rețelei anchetatorilor de opinie, prin instruirea adecvată a acestora, prin utilizarea unor tehnici adecvate de corecție, a estimării și minimizării erorilor.

De ce folosim sondajele de opinie?

Având în vedere că un sondaj corect realizat reproduce structura populației investigate pe principalele ei caracteristici, de aici vor rezulta avantajele utilizării sondajelor de opinie. Sondajele se utilizează atunci când cercetarea statistică exhaustivă (cazul recensământului populației) implică cheltuieli

foarte mari și consum uriaș de eforturi umane și de timp. De asemenea, trebuie avut în vedere că sunt situații când un sondaj statistic bine efectuat, dă rezultate mai bune decât investigarea întregii populații din universul de eșantionare, și aceasta deoarece:

- programul de cercetare prin sondaj cuprinde, de regulă, mai multe caracteristici decât o cercetare exhaustivă a populației;
- anchetatorii de teren utilizați într-un sondaj de opinie sunt superior instruiți față de situația unui recensământ, când aceștia sunt improvizați și cu un instructaj sumar.

După cum am văzut, recensământul este studiul exhaustiv al întregii populații din universul de eșantionare.

Eșantionul care urmează a fi extras trebuie să îndeplinească o condiție de bază, potrivit căreia concluziile obținute din studiul eșantionului să se poată extinde asupra întregii populații. Acest proces se va numi **inferență statistică**.

Capacitatea unui eșantion de a reproduce cât mai fidel structurile și caracteristicile populației din care a fost extras, va fi numită **reprezentativitatea** eșantionului.

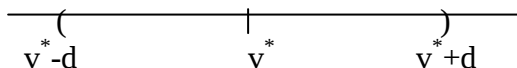
Gradul de reprezentativitate al unui eșantion este măsurat de două mărimi:

- eroarea maximă, notată prin d , și care exprimă diferența cea mai mare pe care o acceptăm între o valoare v^* , dată de eșantion și valoarea teoretică v (dată de populația totală).

Așadar, $d = \max |v^* - v|$

- Mărimea P – numită **nivel de încredere**, care arată ce șanse sunt ca eroarea comisă să nu depășească eroarea maximă d .

În acest fel, se construiește un interval $(v^* - d, v^* + d)$ numit **interval de încredere**



Dacă pentru o anumită caracteristică, valoarea teoretică v se găsește în interiorul acestui interval, $v \in (v^* - d, v^* + d)$ atunci, pentru această caracteristică a populației, eroarea maximă admisă nu este atinsă, iar eșantionul este valid din acest punct de vedere.

Așadar, reprezentativitatea unui eșantion este caracterizată de cuplul (d, P) .

Cu toate acestea, nu putem vorbi de reprezentativitatea a întregului eșantion ci de o reprezentativitate pe fiecare caracteristică în parte.

Reprezentativitatea este o noțiune relativă, în sensul că un eșantion este mai reprezentativ sau mai puțin reprezentativ decât altul.

Compararea se face în felul următor: dacă la un nivel de probabilitate dat, pentru o aceeași caracteristică, eroarea d este mai mică în primul eșantion, atunci acest eșantion est mai reprezentativ.

De asemenea, dacă la o aceeași eroare, nivelul de încredere P este mai ridicat, din nou putem spune că acest eșantion este mai reprezentativ.

Pentru nivelul de încredere P se acceptă valoarea minimă: 0,95 (ceea ce înseamnă că șansele de a greși să nu fie mai mari de 0,05 sau 5%)

Pentru d se acceptă o valoare mai mică decât 3% în sondajele de opinie.

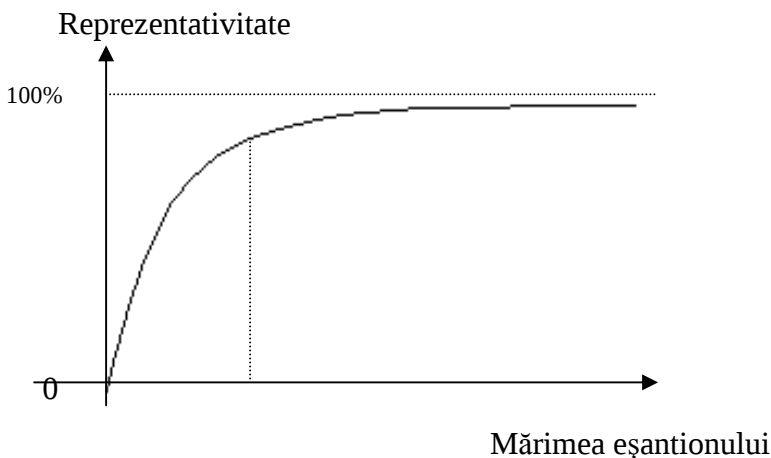
Mărimile d și P nu sunt independente.

Gradul de reprezentativitate al unui eșantion depinde de:

- caracteristicile populației
- mărimea eșantionului
- procedura de eșantionare folosită.

Pentru caracterizarea nivelului de omogenitate al populației se utilizează abaterea standard, care măsoară nivelul de dispersie al indivizilor în jurul mediei.

Să mai reținem că reprezentativitatea crește odată cu creșterea volumului eșantionului, pe anumite porțiuni. Acest lucru este pus în evidență de următorul grafic:



Relația dintre volumul eșantionului și reprezentativitatea sa.

Se constată că peste o anumită limită, creșterea în volum a eșantionului nu mai este justificată de ameliorarea reprezentativității.

Mărimea populației nu intervine în mărimea și reprezentativitatea eșantionului.

Proceduri de eşantionare

După modul cum sunt concepute, eşantioanele sunt de două feluri:

- aleatoare (sau probabilistice)
- nealeatoare

Eşantioanele nealeatoare pot fi dirijate şi mixte.

O procedură de eşantionare se va numi **aleatoare** atunci când fiecare individ din populaţie are o şansă reală, calculabilă şi nenulă de a fi ales în eşantion.

Orice altă procedură este neprobabilistă. Pentru a fi siguri că fiecare individ al populaţiei are şanse de a fi ales în eşantion, este necesară o foarte bună cunoaştere a structurii populaţiei (a universului de eşantionare). Asta înseamnă existenţa unor liste ale populaţiei, adică al unui **cadru de eşantionare**, care să permită accesul la fiecare individ al populaţiei.

Tipuri de eşantionări.

1. **Eşantionarea simplu aleatoare** are la bază principiul loteriei sau al tragerii la sorţi.

Practic, se vor folosi tabelele de numere aleatoare.

2. **Eşantionarea prin stratificare** se efectuează în populaţii neomogene, alcătuite din subpopulaţii omogene sau straturi.

Criteriile de determinare a structurilor sunt:

- **calitative** (geografic: județe, zone, localități; salariat / nesalariat; mediu de proveniență)
- **cantitative** (numărul de membri ai unei familii, cifră de afaceri, dimensiunea localității, etc).

După delimitarea celor s straturi: $N_1, N_2, \dots, N_s$ se extrag în mod simplu aleator s – subeșantioane de mărimi: $n_1, n_2, \dots, n_s$. Aceste volume se extrag fiecare din stratul corespunzător, și sunt proporționale cu mărimea stratului:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_s}{N_s}$$

Se poate arăta că dintre două eșantioane de volum egal, cel realizat prin stratificare are o reprezentativitate mai mare decât cel obținut prin tehnica simplă aleatoare.

3. **Eșantionarea multistadială** (sau grupală).

Aceasta presupune o grupare a populației pe arii geografice, culturale sau județe. În cadrul acestor arii se selectează un număr de localități, în cadrul acestora se selectează un număr de străzi, etc.

Un eșantion multistadial este mai puțin reprezentativ, la volume egale, decât unul simplu aleator, dar comportă un cost mai scăzut.

4. **Eșantionarea multifazică** constă în alegerea unui eșantion mare, la nivelul căruia se aplică un instrument de

cercetare mai simplu; acest eșantion se supune unor operații succesive de eșantionare obținându-se straturi din ce în ce mai mici, cărora li se aplică metode mai elaborate.

5. Eșantionarea pe cote (nealeatoare).

Aceasta presupune gruparea populației după câteva caracteristici și apoi se determină mărimea subeșantioanelor. Structura generală a populației se grupează de regulă după caracteristicile de bază (sex, grupe de vârstă, categorii socio-profesionale), zone urbane (rurale, județe, etc)

În sondajul pe cote, care se aseamănă cu cel stratificat, fiecărui operator îi este repartizat un număr de subiecți, aleși după câteva criterii (câte persoane de sex masculin și câte de sex feminin trebuie luate, câte din fiecare grupă de vârstă, etc).

Modul cum este ales fiecare subiect în parte este lăsat adesea la latitudinea operatorului de teren.

6. Eșantioane fixe (panel). Acestea, odată fixate, sunt supuse unor investigații repetate cu același chestionar. El urmărește schimbările care se petrec în cadrul populației. Prezintă dezavantajul uzurii morale a eșantionului.

Concluzii.

Metoda sondajului de opinie prezintă și o serie de dezavantaje, dintre care cel mai important este acela că

sondajele, de cele mai multe ori, nu surprind schimbările care se petrec în evoluția unui fenomen social.

Sondajele reflectă, de regulă, o situație de moment, ele fiind de fapt o radiografiere a fenomenului studiat la un moment dat.

În anii din urmă, sondajele de opinie au devenit o practică frecventă în viața social politică a țării. Ele măsoară interesul populației pentru viața social-politică, pentru anumite decizii de interes local sau național.

Putem deduce de aici că politica unui guvern, a unei formațiuni politice, a unei întreprinderi, a conducerilor sindicale, se reglează permanent prin intermediul sondajelor care exprimă de fapt interesul unei colectivități.

Să reținem că în cercetările sociologice predomină următoarele tipuri de sondaje: pe cote, aleator și mixt.

Sondajul mixt face un compromis, efectuându-se o cotare sumară (pe 2-3 caracteristici, de ex. zone tradiționale, dimensiunea localităților, etc.) iar în cadrul straturilor alegerea se efectuează aleator. Însăși Hubert Blalock, celebrul metodolog american, subliniază avantajul din punct de vedere practic, al acestei metode.

Tipuri de erori în sondajele statistice

Pe parcursul elaborării și aplicării sondajului statistic se pot introduce o serie de erori, unele semnificative, altele mai puțin.

E bine, însă, de știut, că anumite tipuri de erori pot produce abateri importante ale rezultatului de la realitate.

Erorile se sondaj sunt de două feluri:

- erori de înregistrare
- erori de reprezentativitate

Erorile de înregistrare se pot ține ușor sub control atunci când pentru realizarea anchetei de teren se apelează la personal calificat și experimentat.

Erorile de reprezentativitate pot fi și ele de două feluri:

- erori sistematice
- erori întâmplătoare

Erorile sistematice apar atunci când nu se respectă principiile de bază ale teoriei eșantionării.

Să dăm câteva exemple în care apar erori sistematice de reprezentativitate: afectarea caracterului aleator al sondajului prin selectarea la întâmplare a elementelor statistice, fără a respecta o metodologie anume; intelectualizarea sondajului care are loc atunci când chestionarul de opinie cuprinde întrebări dificile, sofisticat formulate (care induc așa numita “spirală a tăcerii”), fie din comoditate, unii operatori de anchetă

aleg cu precădere subiecți mai școliți și deci mai comozi în desfășurarea anchetei, și aceasta în detrimentul celorlalte categorii; în acest fel sunt afectate proporțiile straturilor, și deci reprezentativitatea eșantionului.

Lipsa conștiinciozității operatorilor de teren și o mare cantitate a nonrăspunsurilor constituie cauzele principale ale erorilor sistematice de reprezentativitate.

Erorile întâmplătoare sau aleatoare de selecție apar în procesul derulării sondajului chiar și atunci când sunt respectate regulile metodologice. Acest tip de eroare provine din structura metodei de eșantionare. Aceste tipuri de erori sunt cunoscute, calculate anterior și se pot ține sub control.

Capitolul IX

CHESTIONARUL DE OPINIE.

ELEMENTE PRIVIND

PROIECTAREA CHESTIONARULUI.

Chestionarul constituie principalul instrument de culegere a datelor prin metoda sondajului statistic de opinie. El nu constituie o simplă înșiruire de întrebări, fără legătură între ele și mai ales fără o anumită logică. Construcția unui chestionar de opinie a fost îmbunătățită permanent, datorită experienței practice acumulate, dar sistematizarea acestuia s-a produs odată cu axiomatizarea chestionarului de către Claude Picard și apoi pe baza analizei informaționale fundamentată de școala românească de statistică după ideile acad. Octav Onicescu. Câteva lucruri se impun a fi menționate.

Se spune adesea că “un sondaj nu poate fi mai bun decât chestionarul său”, adică de modul cum sunt formulate întrebările, de ordinea și complexitatea lor.

Se știe că la întrebările cu mai multe înțelesuri se vor primi răspunsuri echivoce, neconcludente. De asemenea, un chestionar cu multe întrebări, și acestea cu formulări greoaie, produc disconfort atât operatorului, care adesea îl va trata

superficial, dar și respondentului, care se va plictisi, va da semne de nervozitate, nu va mai fi atent la întrebări.

În acest caz, el va răspunde monosilabic sau deloc, ducând la creșterea numărului de non-răspunsuri.

Proiectarea unui chestionar de opinie trebuie să înceapă cu specificarea problemei de cercetat. Problemele sociale au în general un grad mare de complexitate care impun descompunerea lor pe mai multe dimensiuni. Aceste dimensiuni trebuie transformate în indicatori, adică în modalități de stabilire a prezenței sau absenței unei caracteristici, a intensității acesteia.

Fiecare întrebare din chestionar va reprezenta un indicator.

Selectarea întrebărilor care urmează să fie incluse în chestionar, presupune existența unor ipoteze sau chiar a unei teorii privind fenomenul social ce urmează să fie cercetat.

O atenție deosebită trebuie acordată construirii scalelor de răspunsuri care implică atât posibilitatea de ierarhizare cât și cea de măsurare a atitudinilor, avându-se în vedere că se pornește de la opinii pentru a se ajunge la atitudini.

Este indicat să se respecte un număr de reguli care urmăresc să dea întrebărilor și răspunsurilor o formă coerentă care să permită valorificarea corectă a acestora.

Astfel:

- întrebările trebuie să se refere la opinii și nu la fapte;
- întrebarea trebuie să fie la obiect, scurtă și pusă într-un limbaj adecvat, pentru a fi accesibilă;
- întrebarea trebuie să prevadă toate răspunsurile posibile iar anchetatorul de teren să nu favorizeze din ton sau nuanță vreunul din răspunsuri.
- întrebările trebuie puse cu tact și un anume menajament față de subiect pentru a nu-i provoca reacții nedorite.

Tipuri de întrebări

- a) După conținutul lor, întrebările sunt: factuale, de opinie, de cunoaștere

Întrebările factuale privesc aspecte de comportament ale indivizilor anchetați sau ale altora care vin în contact cu ele. Aceste informații sunt, teoretic, verificabile (ce reviste, ziare citește, ce emisiuni TV a urmărit, etc.)

Întrebările de opinie vizează părerile, atitudinile, credințele, atașamentul față de anumite valori, etc. Aceste informații nu pot fi obținute direct prin ale metode.

Întrebările de cunoaștere evidențiază preocupările intelectuale ale indivizilor. Ele pot fi utilizate și ca întrebări de

control (Astfel de întrebări nu aduc un plus de informație, ele verifică răspunsurile de la alte întrebări anterioare).

b) După forma de înregistrare a răspunsurilor, avem:

- întrebări închise
- întrebări deschise
- întrebări cu posibilități multiple de răspuns (se pot alege 2 sau mai multe variante de răspuns).

Analiza non-răspunsurilor

Prin non-răspunsuri înțelegem atât lipsa răspunsurilor la unele întrebări în cadrul aceluiași chestionar, cât și lipsa răspunsului la întregul chestionar.

Apariția non-răspunsurilor în cadrul aceluiași chestionar poate însemna că anumite întrebări sunt dificile, că ele pot deranja anumite segmente ale populației, că nu prevăd toate variantele de răspuns. Unele persoane nu răspund de teamă, din necunoașterea răspunsului sau alte situații.

În orice caz, procente crescute de non-răspunsuri perturbă rezultatul general al sondajului. Toate aceste lucruri trebuie avute în vedere încă din faza de concepere și proiectare a cercetării și a chestionarului.

A doua situație în care chestionare întregi rămân necompletate, este generată de următoarele situații: refuzul de a

coopera, absența de la domiciliu a respondentului în momentul desfășurării anchetei, schimbări de adrese și neactualizate pe listele de eșantionare, cazuri de infirmități, alte unități statistice care au dispărut din baza de sondaj.

Deși statistica oferă soluții care remediază anumite situații, totuși este important să se identifice cauzele care provoacă non-răspunsurile.

Analiza scalară

Analiza scalară sau analiza ierarhică presupune atât posibilitatea de eșantionare cât și măsurarea atitudinilor.

În științele sociale întâlnim patru feluri de scale:

1) **Scale nominale** care presupun o enumerare de posibilități.

Exemplu: locul de proveniență

Termenii acestei scale nu pot fi comparați.

2) **Scale ordinale** care permit o oarecare măsurare a distanței dintre posibilia termeni ai scalei.

Exemplu: note primite la examen, trepte de învățământ, etc, în care fiecare treaptă este superioară celei anterioare.

3) **Scale cu intervale** care permit măsurarea distanțelor dintre diferitele trepte și ranguri .

Exemplul cel mai elocvent este cel al măsurării cronologice.

4) **Scale proporționale** care exprimă posibilitatea unui raport între două poziții de pe o scală.

De exemplu putem spune că o persoană care are 40 ani este de două ori mai în vârstă decât una de 20 ani.

Capitolul X

TESTE DE SEMNIFICATIE

Problema semnificației unor mărimi și mai ales a semnificației diferenței dintre două mărimi se pune atunci când se compară două valori, dintre care cel puțin una provine dintr-o cercetare concretă.

Adoptarea unui plan de selecție la o populație stratificată se sprijină pe un șir de ipoteze: omogenitatea straturilor, volumele eșantioanelor, etc. Un alt plan de selecție presupune alte ipoteze, alte estimatii. În această situație, trebuie analizat căror estimatii le acordăm mai multă încredere, și a vedea în același timp, în ce măsură diferența dintre ele este sau nu semnificativă.

Dacă avem în vedere o anumită caracteristică a variabilei teoretice X și dacă valori estimative λ_1^* și λ_2^* ale valorii teoretice λ , atunci media teoretică a variabilei diferență trebuie să se anuleze, adică

$$M(\lambda_1^* - \lambda_2^*) = 0$$

Acest lucru a condus la introducerea metodei de cercetare denumită **ipoteza nulă**.

Aplicarea ipotezei nule în sensul de a constata dacă această relație este sau nu îndeplinită, poate fi făcută în două moduri:

- prin utilizarea intervalelor de încredere
- prin utilizarea unor criterii, numite teste ale ipotezei nule, bazate pe ipoteza că variabila $d = \lambda_1^* - \lambda_2^*$ are media nulă.

Acceptarea ipotezei nule admite următoarele alternative:

1. Ipoteza făcută este adevărată și urmează să fie acceptată;
2. Ipoteza făcută este falsă și s-a comis o eroare că ea a fost acceptată.

Conceptul de prag de semnificație este o probabilitate și măsoară riscul de a greși atunci când se ia o astfel de hotărâre.

Există teste de semnificație care se aplică eșantioanelor mari, precum: testul Z, testul χ^2 și teste de semnificație pentru eșantioane mici: testul t (al lui Student), testul F (Fisher-Snedecor)

În cele ce urmează, vom analiza testul χ^2 .

Se pune problema de a testa dacă structura eșantionului se abate semnificativ de la o structură standard, după o caracteristică.

Să luăm un exemplu. Să presupunem un eșantion de 1000 de nașteri cu următoarea distribuție:

Anotimp	Primăvara	Vara	Toamna	Iarna	Total
Număr	240	270	280	210	1000
Procent	24%	27%	28%	21%	100%

Se constată că subeșantioanele nu sunt uniform distribuite pe anotimpuri.

Se pune problema dacă această serie diferă semnificativ de cea în care respectivele proporții ar fi identice: 25%

Testul χ^2 se aplică astfel:

Notăm prin $k_1, k_2, \dots, k_s$ un șir de frecvențe obținute pe un eșantion și cu $m_1, m_2, \dots, m_s$ frecvențele teoretice corespunzătoare. Formula este:

$$\chi^2 = \sum_{i=1}^s \frac{(k_i - m_i)^2}{m_i}$$

De o deosebită importanță în utilizarea testului χ^2 este stabilirea numărului **gradelor de libertate**. Acesta se calculează astfel:

$v = s - 1$ în cazul unui tabel unidimensional cu s celule;

$v = (s-1)(t-1)$ în cazul unui tabel bidimensional cu $s \times t$ celule.

În cazul exemplului anterior, avem:

Șirul frecvențelor empirice: 240; 270; 280; 210

Șirul frecvențelor teoretice: 250; 250; 250; 250

$$v = 4 - 1 = 3$$

Din tabele găsim valorile critice:

7,82 pentru pragul de semnificație $p=0,05$

9,84 pentru pragul de semnificație $p=0,02$

11,35 pentru pragul de semnificație $p=0,01$

$$\begin{aligned}\chi^2 &= \frac{(240 - 250)^2}{250} + \frac{(270 - 250)^2}{250} + \frac{(280 - 250)^2}{250} + \frac{(210 - 250)^2}{250} = \\ &= \frac{100 + 400 + 900 + 1600}{250} = \frac{3000}{250} = 12\end{aligned}$$

Cum $12 > 7,82$ (valoarea critică) rezultă că diferența este semnificativă.

Altfel spus, **ipoteza nulă** este respinsă cu o probabilitate de 0,95.

Capitolul XI

MODELE DE ANALIZĂ

A CARACTERISTICILOR CALITATIVE

Def. Numim **caracteristică calitativă** o anumită proprietate de care se bucură elementele populației statistice studiate, și care are în vedere natura lor specifică.

Studiul caracteristicilor calitative este legat de **asocierea** sau neasocierea acestora, sau la definirea **interdependenței** dintre ele.

Caracterul cantitativ rezultă din numărarea și ordonarea unităților populației, care au sau nu au proprietatea (P) considerată.

Astfel, dacă de exemplu, populația statistică este formată dintr-un număr de produse a căror calitate trebuie testată, rezultatul testării are două valori: acceptat sau neacceptat. Sau dacă populația statistică este populația unei localități, care urmează să fie (investigată cu ajutorul unui chestionar de opinie) anulată în raport cu anumită problemă, răspunsul se va da prin Da sau Nu. Astfel de situații arată că avem de-a face cu caracteristici calitative care capătă forma unei variabile aleatoare dihotomice (cu două valori). Numărul sau proporția

elementelor observate definesc în acest fel caracteristica calitativă corespunzătoare.

De multe ori, metodele statistice utilizate în cazul analizei cantitative, pentru obținerea de informații statistice, se pot transfera și în cazul analizei caracteristicilor calitative. Prezența sau absența unei caracteristici calitative poate fi privită ca două evenimente aleatoare incompatibile cărora li se pot atașa valori 0 și 1, și a căror distribuție urmează, în linii mari, modelul legii binomiale.

Studiul caracteristicilor calitative începe cu operația de grupare a elementelor, care înseamnă o separare a populației statistice în grupe omogene de elemente. Operația de grupare conduce la crearea de subgrupe ale populației statistice pe criteriul alternativ – dihotomic: subgrupe cu proprietatea (P) și fără proprietatea (P) notate prin (A) și ($\bar{A}$). Astfel că dacă N este populația statistică studiată, atunci $(A) + (\bar{A}) = N$. Partiția populației statistice poate continua prin grupări de ordinul doi după o nouă caracteristică.

O nouă astfel de partiție după caracteristica B conduce la constituirea grupelor: AB, $A\bar{B}$, $\bar{A}B$, $\bar{A}\bar{B}$. Are loc relația:

$$(AB) + (A\bar{B}) = (A), \quad (AB) + (\bar{A}B) = (B)$$

și partiția poate continua după noi caracteristici.

Astfel, pentru o partiție după trei caracteristici A, B, C apar subgrupele de ordinul 3:

ABC , $AB\bar{C}$, $A\bar{B}C$, $A\bar{B}\bar{C}$, $\bar{A}BC$, $\bar{A}\bar{B}C$, $\bar{A}\bar{B}\bar{C}$, $\bar{A}\bar{B}\bar{C}$ și odată cu ele au loc relațiile:

$(ABC) + (AB\bar{C}) = (AB)$; $(ABC) + (A\bar{B}C) = (AC)$, etc.

Exemplu. Să presupunem că personalul angajat al unei firme a fost clasificat după caracteristicile:

A: bărbat

B: vârsta de până la 30 de ani

C: studii superioare (universitare)

Înregistrarea personalului după aceste patru caracteristici s-a realizat în următorul tabel:

Grupele finale	Frecvența absolută
$A B C$	125
$A B \bar{C}$	50
$A \bar{B} C$	25
$\bar{A} B C$	30
$A \bar{B} \bar{C}$	20
$\bar{A} \bar{B} C$	55
$\bar{A} B \bar{C}$	45
$\bar{A} \bar{B} \bar{C}$	150
TOTAL	500

Tabel 1

Pe baza datelor centralizate în acest tabel, vom putea determina frecvențele grupelor de ordin inferior, caracterizate

prin prezența numai a caracteristicilor A, B, C. Să remarcăm mai întâi care sunt caracteisticile complementare. Astfel:

$\bar{A}$: femeie

$\bar{B}$: vârsta de peste 30 ani

$\bar{C}$: studii preuniversitare

Astfel, grupa de ordinul 0 este chiar personalul angajat al firmei: 500 persoane.

Grupa (AB) care înseamnă “bărbați cu vârsta de până la 30 ani”.

$$(AB) = (ABC) + (AB\bar{C}) = 125 + 50 = 175$$

$$(AC) = (ABC) + (A\bar{B}C) = 125 + 25 = 150$$

$$(BC) = (ABC) + (\bar{A}BC) = 125 + 30 = 155$$

$$(A) = (AB) + (A\bar{B}) = (AB) + (A\bar{B}C) + (A\bar{B}\bar{C}) = 175 + 25 + 20 = 220$$

$$(B) = (AB) + (\bar{A}B) = (AB) + (\bar{A}BC) + (\bar{A}B\bar{C}) = 175 + 30 + 55 = 260$$

$$(C) = (AC) + (\bar{A}C) = (AC) + (\bar{A}BC) + (\bar{A}\bar{B}C) = 150 + 30 + 55 = 235$$

Observație. Frecvențele de grupă și analizele statistice realizate pe baza acestora, au proprietatea de concordanță, adică de a nu conduce la contradicții.

Se poate arăta că **o serie de frecvențe de grupă are proprietatea de concordanță dacă și numai dacă frecvențele finale sunt pozitive.**

Astfel, dacă considerăm populația statistică N dihotomizată după două caracteristici A și B cu grupele de frecvențe finale (AB) , $(\bar{A}B)$, $(A\bar{B})$ și $(\bar{A}\bar{B})$, se obțin inegalitățile:

$$(AB) \geq 0$$

$$(AB) \geq (A) + (B) - N$$

$$(AB) \leq (A)$$

$$(AB) \leq (B)$$

De asemenea, unei repartiții dihotomice de trei caracteristici A, B, C , se obțin condițiile:

$$(ABC) \geq 0$$

$$(ABC) \geq (AB) + (AC) - (A)$$

$$(ABC) \geq (AB) + (BC) - (B)$$

$$(ABC) \geq (AC) + (BC) - (C)$$

$$(ABC) \leq (AB)$$

$$(ABC) \leq (AC)$$

$$(ABC) \leq (BC)$$

$$(ABC) \leq (AB) + (AC) + (BC) - (A) - (B) - (C) + N$$

Aceste inegalități au drept consecință următoarele:

$$(AB) + (AC) + (BC) \geq (A) + (B) + (C) - N$$

$$(AB) + (AC) \leq (BC) + (A)$$

$$(AB) + (BC) \leq (AC) + (B)$$

$$(AC) + (BC) \leq (AB) + (C)$$

și în plus:

$$(AB) \geq 0, (AC) \geq 0, (BC) \geq 0$$

$$(A) \geq (AC), (A) \geq (AB)$$

$$(B) \geq (AB); (B) \geq (BC)$$

$$(C) \geq (AC); (C) \geq (BC)$$

$$(BC) \geq (B) + (C) - N$$

$$(AC) \geq (A) + (C) - N$$

$$(AB) \geq (A) + (B) - N$$

Să consideră următorul exemplu. La o sesiune de examene, dintr-o sută de studenți care au susținut examenele la disciplinele analiză, algebră și informatică, 75 dintre aceștia au promovat examenul de analiză, 80 au promovat examenul de algebră și 65 studenți au promovat examenul de informatică. Care este numărul minim și cel maxim de integraliști?

Într-adevăr, conform inegalităților

$$(ABC) \geq (AB) + (AC) - (A) \geq (A) + (B) - N + (A) + (C) - N - (A)$$

deci:

$$(ABC) \geq (A) + (B) + (C) - 2N = 75 + 80 + 65 - 200 = 20$$

Așadar, numărul minim posibil de studenți integraliști este de 20.

În fine, numărul maxim posibil este dat de formula:

$$(ABC) \leq (AB) + (AC) + (BC) - (A) - (B) - (C) + N \leq \min\{(A), (B)\} + \min\{(A), (C)\} + \min\{(B), (C)\} - (A) - (B) - (C) + N = 75 + 65 + 65 - 75 - 80 - 65 + 100 = 85$$

Pentru a vedea în ce condiții este atinsă limita inferioară de 20 de integraliști, trebuie ca $(ABC) = (A) + (B) + (C) - 2N$

Dar

$$N = (ABC) + (\bar{A}BC) + (A\bar{B}C) + (AB\bar{C}) + (A\bar{B}\bar{C}) + (\bar{A}B\bar{C}) + (\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C}) \text{ iar}$$

$$(A) = (ABC) + (A\bar{B}C) + (AB\bar{C}) + (A\bar{B}\bar{C}), \text{ etc.}$$

Egalând, obținem:

$$\begin{aligned} (ABC) &= (ABC) + (A\bar{B}C) + (AB\bar{C}) + (A\bar{B}\bar{C}) + (ABC) + (\bar{A}BC) + (AB\bar{C}) + (\bar{A}B\bar{C}) + (ABC) + (\bar{A}BC) + (A\bar{B}C) + (\bar{A}\bar{B}C) \\ &\quad - 2(ABC) - 2(\bar{A}BC) - 2(A\bar{B}C) - 2(AB\bar{C}) - 2(A\bar{B}\bar{C}) - 2(\bar{A}B\bar{C}) - 2(\bar{A}\bar{B}C) - 2(\bar{A}\bar{B}\bar{C}) \end{aligned}$$

De aici rezultă:

$$(A\bar{B}\bar{C}) + (\bar{A}B\bar{C}) + (\bar{A}\bar{B}C) + 2(\bar{A}\bar{B}\bar{C}) = 0$$

Acest lucru nu este posibil decât dacă fiecare termen este nul, adică:

$$(A\bar{B}\bar{C}) = 0, (\bar{A}B\bar{C}) = 0, (\bar{A}\bar{B}C) = 0 \text{ și } (\bar{A}\bar{B}\bar{C}) = 0$$

Acest lucru înseamnă că limita inferioară este atinsă, adică sunt exact 20 de integraliști, atunci când nu există nici un student care să fi pierdut mai mult de un examen.

De asemenea, putem vedea în ce condiții est atinsă limita superioară de studenți integraliști. Acest lucru se întâmplă atunci când anumite inegalități devin egalități, de exemplu, ar urma să aibă loc relația:

$$(ABC) = (AB) + (AC) + (BC) - (A) - (B) - (C) + N \quad (*)$$

$$\text{Dar } (AB) = (ABC) + (A\bar{B}\bar{C}), (AC) = (ABC) + (A\bar{B}C) \text{ și} \\ (BC) = (ABC) + (\bar{A}BC), \text{ iar}$$

$$(A) = (ABC) + (A\bar{B}\bar{C}) + (A\bar{B}C), (A\bar{B}\bar{C})$$

$$(B) = (ABC) + (\bar{A}BC) + (A\bar{B}C) + (\bar{A}\bar{B}C)$$

$$(C) = (ABC) + (\bar{A}BC) + (A\bar{B}C) + (\bar{A}\bar{B}C)$$

De asemenea,

$$N = (ABC) + (\bar{A}BC) + (A\bar{B}C) + (A\bar{B}\bar{C}) + (\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C}).$$

Înlocuind toate aceste relații în relația (*) obținem:

$$(\bar{A}\bar{B}\bar{C}) = 0$$

ceea ce înseamnă că limita superioară va fi atinsă atunci când toți studenții promovează cel puțin un examen.

Dihotomizarea populației statistice după caracteristicile de grupă ne permit să interpretăm frecvențele de grupă de orice ordin ca o probabilitate. Astfel, raportul $\frac{A}{N}$ va însemna

probabilitatea de apariție la o extracție (alegere) a unui element statistic cu proprietatea A.

Caracteristici statistice independente

Vom spune că două caracteristici A și B sunt independente dacă nu au nici o legătură între ele.

Drept criteriu de independență pentru caracteisticile A și B vom lua îndeplinirea relației:

$$\frac{(AB)}{(B)} = \frac{(\overline{AB})}{(\overline{B})} \quad (1)$$

Relația (1) este necesară pentru independența variabilelor A și B, nu și suficientă. Se poate arăta că dacă relația (1) are loc, atunci au loc și relațiile:

$$\begin{aligned} \frac{(AB)}{(A)} &= \frac{(\overline{AB})}{(\overline{A})} \\ \frac{(\overline{AB})}{(B)} &= \frac{(\overline{AB})}{(\overline{A})} \\ \frac{(\overline{AB})}{(A)} &= \frac{(\overline{AB})}{(\overline{A})} \end{aligned} \quad (2)$$

Să constatăm că și relația (1) poate căpăta o nouă formă. Astfel,

$$\frac{(AB)}{(B)} = \frac{(\overline{AB})}{(\overline{B})} = \frac{(AB) + (\overline{AB})}{(B) + (\overline{B})} = \frac{(A)}{N}. \text{ De aici deducem}$$

$$(AB) = \frac{(A)(B)}{N} \text{ sau:}$$

$$\frac{(AB)}{N} = \frac{(A)}{N} \cdot \frac{(B)}{N} \quad (3)$$

condiție care se transpune în următorul criteriu:

Criteriu. O condiție necesară ca două caracteristici A și B să fie independente este ca frecvența elementelor AB să fie egală cu produsul frecvențelor elementelor A și elementelor B.

Observație. Relația (3) nu este singura care caracterizează independența caracteristicilor A și B. Au loc încă trei relații similare și anume:

$$\begin{aligned} \frac{(\overline{AB})}{N} &= \frac{(A)}{N} \cdot \frac{\overline{B}}{N} \\ \frac{(\overline{AB})}{N} &= \frac{(\overline{A})}{N} \cdot \frac{(B)}{N} \\ \frac{(\overline{AB})}{N} &= \frac{(\overline{A})}{N} \cdot \frac{(\overline{B})}{N} \end{aligned} \quad (4)$$

În fine, un al treilea tip de criteriu care caracterizează independența caracteristicilor A și B se deduce din cele de mai sus și are forma:

$$(\overline{A} B) \cdot (A \overline{B}) = (AB) \cdot (\overline{A} \overline{B}) \quad (5)$$

Se poate arăta că aceste criterii reiese sub forme diferite, sunt și suficiente.

Să luăm un exemplu.

Într-o populație statistică de N subiecți, se fac înregistrări ale gupelor de ordinul doi ale caracteristicilor A și B , obținându-se frecvențele de grupă următoare:

$$(AB) = 125; (\bar{A} B) = 75; (A \bar{B}) = 250; (\bar{A} \bar{B}) = 225.$$

Putem, afirma că cele două caracteristici sunt îndeplinite?

Pentru a răspunde, să utilizăm criteriul dat de formula (5):

$$(\bar{A} B) \cdot (A \bar{B}) = 75 \cdot 250 = 18.750, \text{ iar}$$

$$(AB) \cdot (\bar{A} \bar{B}) = 125 \cdot 225 = 28.125$$

Cum $(\bar{A} B) \cdot (A \bar{B}) \neq (AB) \cdot (\bar{A} \bar{B})$, rezultă că cele două caracteristici nu sunt independente.

Neîndeplinirea unuia din criterii dovedește faptul că între caracteristicile A și B există o anumită legătură. Faptul că are loc una din inegalitățile

$$\frac{(AB)}{N} > \frac{(A)}{N} \cdot \frac{(B)}{N} \text{ sau } \frac{(AB)}{N} < \frac{(A)}{N} \cdot \frac{(B)}{N}$$

indică o anumită **asociere** între caracteristicile A și B care în cazul primei inegalități vom spune că este o asociere de tip pozitiv, iar în cazul celei de-a doua, de tip negativ.

În termeni de mulțimi, dacă $A \subset B$ (sau $B \subset A$) spumem că avem de-a face cu o **asociere completă**. În cazul în care are loc o asociere completă, cu $A \subset B$, atunci $(AB) = A$ iar dacă $B \subset A$ rezultă $(AB) = (B)$.

Să considerăm următorul exemplu. În tabelul de mai jos sunt înscrise frecvențele de grupă ale unei populații în care s-au urmărit caracteristicile:

A = băutor – consumator de băuturi tari și B = bolnav de ficar

Caracteristici	B (bolnav de ficat)	$\bar{B}$ (sănătos)	Total
A (băutor)	240	60	300
$\bar{A}$ (nebăutor)	110	590	700
Total	350	650	1000

Tabelul 2

Ne propunem să studiem interdependența dintre cele două caracteristici A și B. Astfel, dorim să vedem în ce măsură consumul de băuturi alcoolice tari se asociază cu afecțiuni ale ficatului.

Pentru aceasta vom calcula:

$$\frac{(AB)}{(A)} = \frac{240}{300} = 80\% \text{ (procentul consumatorilor de băuturi}$$

alcoolice cu afecțiuni ale ficatului, în totalul băutorilor)

$$\frac{(\bar{A}\bar{B})}{(\bar{A})} = \frac{110}{700} = 15,71\% \text{ (procentul celor care nu consumă}$$

alcool dar cu afecțiuni ale ficatului din totalul celor care nu consumă alcool).

$\frac{(\overline{AB})}{(A)} = \frac{60}{300} = 20 \%$ (procentul băutorilor sănătoși în totalul băutorilor)

$\frac{(\overline{\overline{AB}})}{(\overline{\overline{A}})} = \frac{590}{700} = 84,28 \%$ (procentul celor care nu consumă alcool și sunt sănătoși în totalul nebăutorilor).

Comparând $\frac{(\overline{AB})}{(A)}$ cu $\frac{(\overline{\overline{AB}})}{(\overline{\overline{A}})}$ constatăm că are loc inegalitatea:

$$\frac{(\overline{AB})}{(A)} > \frac{(\overline{\overline{AB}})}{(\overline{\overline{A}})}$$

ceea ce dovedește statistic că obiceiul de a consuma băuturi tari și afecțiunile ficatului sunt puternic asociate pozitiv.

Inegalitatea:

$$\frac{(\overline{AB})}{(A)} < \frac{(\overline{\overline{AB}})}{(\overline{\overline{A}})}$$

care arată că există o asociere negativă între consumul de alcool și starea de sănătate a pacientului, vine să confirme concluziile inițiale.

Intensitatea asocierii

Modul de intensitate al asocierii a două caracteristici poate fi măsurat printr-o serie de indicatori. Astfel, se definește **coeficientul de asociere** a caracteristicilor A și B:

$$q_{(AB)} = \frac{(AB)(\overline{AB}) - (\overline{AB})(\overline{AB})}{(AB)(\overline{AB}) + (\overline{AB})(\overline{AB})} \quad (6)$$

în care notăm

$$\delta = \frac{1}{N}[(AB)(\overline{AB}) - (\overline{AB})(\overline{AB})] = (AB) - \frac{(A) \cdot (B)}{N}$$

Coeficientul $q_{(AB)}$ ia valori cuprinse în intervalul $[-1, 1]$. El ia valoarea 0 ($\delta=0$) când caracteristicile A și B sunt independente, ia valoarea +1 când $(\overline{AB})(\overline{AB}) = 0$ și valoarea -1 când $(AB)(\overline{AB}) = 0$

Un alt coeficient de asociere, numit după C.V.Yule și M.C. Kendall, **coeficient de interdependență**, este dat de formula:

$$Y_{AB} = \frac{1 - \sqrt{\frac{(AB)(\overline{AB})}{(AB)(\overline{AB})}}}{1 + \sqrt{\frac{(AB)(\overline{AB})}{(AB)(\overline{AB})}}} \quad (7)$$

Se poate arăta că:

$$q_{(AB)} = \frac{2Y_{AB}}{1 + Y_{AB}^2} \quad (8)$$

Coeficientul Y_{AB} are proprietăți similare cu $q_{(AB)}$.

Pe baza datelor din tabelul 2, să calculăm intensitatea asocierii utilizând expresiile celor doi coeficienți.

$$q_{(AB)} = \frac{(AB) \cdot (\overline{AB}) - (\overline{AB}) \cdot (\overline{AB})}{(AB) \cdot (\overline{AB}) + (\overline{AB}) \cdot (\overline{AB})} = \frac{240 \cdot 590 - 60 \cdot 110}{240 \cdot 590 + 60 \cdot 110} = 0,9109$$

$$Y_{AB} = \frac{1 - \sqrt{\frac{(AB)(\overline{AB})}{(AB)(\overline{AB})}}}{1 + \sqrt{\frac{(AB)(\overline{AB})}{(AB)(\overline{AB})}}} = \frac{1 - \sqrt{\frac{60 \cdot 110}{240 \cdot 590}}}{1 + \sqrt{\frac{60 \cdot 110}{240 \cdot 590}}} = 0,6449$$

Asocieri în populații cu mai multe caracteristici

Dacă avem de-a face cu o populație statistică în care se iau în vedere mai multe caracteristici, pe lângă studiul asocierilor bilaterale este important de stabilit dacă aceste asocieri sunt directe sau prin intermediul unor relații cauzale generate de una sau mai multe caracteristici intermediare.

Aceste supoziții ne conduc la necesitatea introducerii noțiunii de **asociere parțială și totală**, care înseamnă o asociere la nivelul unei subpopulații sau a populației în ansamblul ei.

Vom spune că două caracteristici A și B sunt **asociate pozitiv** în cadrul subpopulației C, dacă are loc inegalitatea:

$$(ABC) > \frac{(AC)(BC)}{(C)} \quad (9)$$

și **negativ asociate**, în caz contrar

$$(ABC) < \frac{(AC)(BC)}{(C)} \quad (10)$$

Asocierea parțială a caracteristicilor A și B în subpopulația (CD) are forma în cazul asocierii parțiale pozitive:

$$(ABCD) > \frac{(ACD)(BCD)}{(CD)} \quad (11)$$

și

$$(ABCD) < \frac{(ACD)(BCD)}{(CD)} \quad (12)$$

în cazul asocierii parțiale negative.

Coeficientu de asociere poate fi generalizat, în cadrul diferitelor subpopulații. Aceștia vor fi **coeficienți ai asocierii parțiale**. Vom utiliza notația (AB,C) care va însemna asocierea dintre caracteristicile A și B în cadrul subpopulației C. Astfel,

$$q_{(ABC)} = \frac{(ABC)(\overline{ABC}) - (\overline{ABC})(\overline{ABC})}{(ABC)(\overline{ABC}) + (\overline{ABC})(\overline{ABC})} \quad (13)$$

În mod analog, vom desemna coeficientul de asociere între caracteristicile A și B în cadrul subpopulației CD astfel:

$$q_{(ABCD)} = \frac{(ABCD)(\overline{ABCD}) - (\overline{ABCD})(\overline{ABCD})}{(ABCD)(\overline{ABCD}) + (\overline{ABCD})(\overline{ABCD})} \quad (14)$$

În mod analog pot fi definiți și **coeficienții de interdependență parțială**. Astfel, coeficientul de interdependență dintre A și B în subpopulația C va avea forma:

$$Y_{ABC} = \frac{1 - \sqrt{\frac{(A\bar{B}C)(\bar{A}BC)}{(AB\bar{C})(\bar{A}\bar{B}C)}}}{1 + \sqrt{\frac{(A\bar{B}C)(\bar{A}BC)}{(AB\bar{C})(\bar{A}\bar{B}C)}}} \quad (15)$$

Să considerăm următorul exemplu. Într-o populație de 1000 de elevi sunt luate în considerare următoarele caracteristici:

A = nota mică la purtare (sub 7)

B = nivel cultural educativ – scăzut

C = frecvență slabă la școală

Observațiile făcute s-au concretizat în următoarele date statistice:

$N = 1000$; $(A) = 96$; $(B) = 145$; $(C) = 80$

$(AB) = 24$; $(AC) = 24$; $(BC) = 51$; $(ABC) = 13$

Pe baza acestor date înregistrate, putem trage concluzia că nota mică la purtare și slaba frecvență se datorează nivelului cultural educativ – scăzut?

Să constatăm că avem de determinat mai mulți indicatori. Să calculăm mai întâi frecvențele finale.

Avem:

$$(A\bar{B}\bar{C}) = (AB) - (ABC) = 24 - 13 = 11$$

$$(\overline{A}BC) = (AC) - (ABC) = 24 - 13 = 11$$

$$(\overline{A}BC) = (BC) - (ABC) = 51 - 13 = 38$$

$$(\overline{A}\overline{B}C) = (\overline{B}C) - (\overline{A}BC) = (C) - (BC) - (\overline{A}BC) = 80 - 51 - 11 = 18$$

$$(\overline{A}\overline{B}C) = (\overline{A}C) - (\overline{A}BC) = (A) - (AC) - (\overline{A}BC) = 96 - 24 - 11 = 61$$

$$(\overline{A}\overline{B}C) = (\overline{B}C) - (\overline{A}BC) = (B) - (BC) - (\overline{A}BC) = 145 - 51 - 11 = 83$$

$$(\overline{A}\overline{B}\overline{C}) = (\overline{B}\overline{C}) - (\overline{A}BC) = N - (C) - (B) + (BC) - (\overline{A}BC) = 1000 - 80 - 145 + 51 - 61 = 765$$

Să calculăm acum frecvențele de grupă de ordinul 2:

$$(\overline{A}\overline{B}) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 18 + 765 = 783$$

$$(\overline{A}C) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 765 + 83 = 848$$

$$(\overline{B}C) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 61 + 765 = 826$$

Deoarece $\overline{A} = N - A$, etc, rezultă că

$$\overline{A} = 1000 - 96 = 904;$$

$$\overline{B} = 1000 - 145 = 855;$$

$$\overline{C} = 1000 - 80 = 920$$

În fine, să calculăm și frecvențele de grupă de ordinul 2.

$$(\overline{A}\overline{B}) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 38 + 83 = 121$$

$$(\overline{A}C) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 11 + 61 = 72$$

$$(\overline{B}C) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 38 + 18 = 56$$

$$(\overline{A}\overline{C}) = (\overline{A}BC) + (\overline{A}\overline{B}C) = 11 + 61 = 72$$

$$(\overline{BC}) = (A\overline{B}\overline{C}) + (\overline{A}\overline{B}\overline{C}) = 11 + 83 = 94$$

$$(\overline{BC}) = (A\overline{B}\overline{C}) + (\overline{A}\overline{B}\overline{C}) = 11 + 18 = 29$$

Pentru a analiza caracteristica “notă mică la purtare”, ca și “nivelul slabei frecvențe” la cursuri, ca urmare a nivelului cultural-educativ scăzut, vom cerceta dacă existența asocierii între A și C este cauzată de asocierile A cu B și C cu D.

Pentru aceasta, vom determina următorii indicatori pe diferite grupe ale populației:

- frecvența slabă la școală raportată la întreaga populație:

$$\frac{(C)}{N} = \frac{80}{1000} = 8\%$$

- frecvența grupului cu nivel cultural – educativ scăzut raportată la întreaga populație:

$$\frac{(B)}{N} = \frac{145}{1000} = 14,5\%$$

- frecvența grupului cu nota mică la purtare în întreaga populație:

$$\frac{(A)}{N} = \frac{96}{1000} = 9,6\%$$

Pentru subpopulația A:

- proporția elementelor subgrupe AC în A:

$$\frac{AC}{A} = \frac{24}{96} = 25\%$$

Pentru subpopulația B:

- proporția subgrupeii cu frecvență slabă raportată la subpopulația B:

$$\frac{(BC)}{(B)} = \frac{51}{145} = 35,17\%$$

- proporția grupei ABC în AB:

$$\frac{(ABC)}{(AB)} = \frac{13}{24} = 54,16\%$$

Pentru subpopulația $\bar{B}$, cu nivel cultural–educativ ridicat:

- proporția subgrupeii cu frecvență slabă în $\bar{B}$:

$$\frac{(\bar{BC})}{(\bar{B})} = \frac{29}{855} = 3,39\%$$

- proporția subgrupeii cu frecvență slabă și note mici la purtare:

$$\frac{(\overline{ABC})}{(\overline{AB})} = \frac{11}{72} = 15,28\%$$

Comparăm (ABC) cu $\frac{(AB)(BC)}{(B)}$, care ne permite să

apreciem asocierea parțială a lui A și C în B:

$$ABC) = 13; \frac{(AB)(BC)}{(B)} = \frac{24 \cdot 51}{145} = 8,44$$

Deci $(ABC) > \frac{(AB)(BC)}{(B)}$, ceea ce înseamnă că A și C sunt

pozitiv asociate în B.

Vom verifica și asocierea parțială a lui A și C în subpopulația $\bar{B}$:

$$(A\bar{B}C) = 11;$$

$$\frac{(A\bar{B}) \cdot (\bar{B}C)}{(\bar{B})} = \frac{72 \cdot 29}{855} = 2,44$$

$$\text{Așadar, } (A\bar{B}C) > \frac{(A\bar{B}) \cdot (\bar{B}C)}{(\bar{B})}$$

ceea ce dovedește o asociere pozitivă a lui A și C în $\bar{B}$.

Am văzut astfel că A și C sunt asociate atât în B cât și în $\bar{B}$. Să vedem relația dintre A și C în ansamblul populației.

Pentru aceasta, vom compara $\frac{(AC)}{N}$ cu $\frac{(A)}{N} \cdot \frac{(C)}{N}$

$$\text{Dar } \frac{(AC)}{N} = \frac{24}{1000} = 0,024$$

$$\frac{(A)}{N} \cdot \frac{(C)}{N} = \frac{96}{1000} \cdot \frac{80}{1000} = 0,096 \cdot 0,08 = 0,00768$$

Așadar,

$$\frac{(AC)}{N} > \frac{(A)}{N} \cdot \frac{(C)}{N}.$$

ceea ce înseamnă că avem o asociere pozitivă a lui A și C și în ansamblul populației.

Toate acestea ne spun că cele două caracteristici A și C nu sunt independente nici la nivelul subpopulațiilor B și $\bar{B}$, și nici la nivelul întregii populații, în care sunt asociate pozitiv.

Capitolul XII

O METODĂ DE ANALIZĂ SCALARĂ ȘI IERARHIZARE

Analiza scalară sau analiza de stabilire a ierarhiilor presupune atât posibilitatea de eşalonare sau evidențierea diferențelor calitative ale atitudinilor printr-o ordonare a acestora, cât și posibilitatea de măsurare, de cuantificare atitudinală, prin care înțelegem punerea în lumină a diferențelor cantitative.

Principalele tipuri de scale întâlnite în domeniul științelor sociale sunt: **scalele nominale**, **scalele ordinale**, **scalele cu intervale** și **scalele proporționale**.

În cazul **scalelor nominale**, fiecare valoare scalară are aceeași pondere, ea constituind de fapt o enumerare a posibilităților. Fiind o scală neparametrică, ea permite clasificarea elementelor studiate în grupe, ale căror elemente diferă prin caracteristica de scalare.

Scalele ordinale permit o anumită clasificare a opiniilor și pot marca unele departajări de poziții (respectiv inferioare, superioare etc.). Un tip des întâlnit de scală ordinală îl constituie nivelul de școlarizare a subiecților.

Construcția unei **scale atitudinale** are la bază principiul de a se porni de la opinii spre a se ajunge la atitudini.

Cuantificarea răspunsurilor și analiza statistică a informației, premergătoare interpretărilor, constituie principala dificultate în construirea scalelor de atitudini. Un procedeu îndeobște folosit și îndelung controversat este acela al scorurilor, prin care se atribuie o valoare fixă, echivalentul unei notări. Suma acestor note dă scorul final, care poate conduce la o clasificare. Principalul punct nevralgic al metodei îl constituie faptul că ele presupun o proporționalitate a scalelor nominale.

După cum se știe, analiza scalară nu poate fi evitată în cercetarea sociologică, iar una din metodele cele mai riguroase de scalare o constituie analiza scalogramă a lui Guttman, care admite că răspunsul favorabil la o întrebare implică adeziunea la toate întrebările anterioare.

Ținându-se cont de principiile potrivit cărora răspunsurile se clasifică, în primul rând, după numărul de poziții pozitive sau negative, prin eliminarea celor aberante, și în al doilea rând prin clasificarea respondenților, luând întâi pe cei care au răspuns pozitiv la toate cele **n** întrebări, apoi pe cei care au răspuns pozitiv la **n-1**, **n-2** ș.a.m.d., se poate ajunge, din aproape în aproape, la scalări iterative perfectibile.

În cele ce urmează propunem o modalitate de construcție scalară, care stabilește ierarhii pe baza întregii cantități de informație pe care o furnizează sistemul cercetat.

Această construcție are la bază conceptul de energie informațională, pe care îl amintim în continuare, și ea va genera o scală pe care o vom numi **scală informațională** (SI). Această scală este standardizată și va putea juca rolul de etalon pentru scalele graduale oferite de frecvențele simple. Ea va putea juca rolul de scală atitudinală, întrucât realizează o clasificare standardizată a opiniilor.

1. Energia informațională a unui sistem cu un număr finit de stări

Dacă un sistem (S) are un număr finit de componente distincte – $s_1, s_2, \dots, s_m$, cu ponderile respective – $p_1, p_2, \dots, p_m$, acestea din urmă sunt caracterizate de condițiile:

$$p_i \geq 0, \sum_{i=1}^m p_i = 1$$

Informația globală a sistemului (S) cu stările (s_i), $i=1, 2, \dots, m$, poate fi exprimată prin **energia sa informațională***, calculată cu ajutorul formulei:

$$E_s = \sum_{j=1}^m p_j^2, \sum_{j=1}^m p_j = 1 \quad (1)$$

Proprietăți. Valoarea energiei informaționale devine **1/m** când toate stările au aceeași pondere, și ia valoarea **1** când una din stări are ponderea egală cu **1**, celelalte fiind nule. Prima situație corespunde unei situații de o totală “nedeterminare” sau la o totală “indiferență reciprocă” între stări diferite, care corespunde unei totale dezorganizări a sistemului. În acest caz are loc:

$$p_1 = p_2 = \dots = p_m = 1/m$$

A doua situație corespunde cazului când una din stări are frecvența **1**.

În acest caz, experimentul dă același rezultat repetat, determinat și unic.

Formula **(1)** nu indică această stare a sistemului.

Energia informațională este cuprinsă între valorile $1/m$ și **1**:

$$1/m \leq E_s \leq 1$$

* Expresia (1) a fost utilizată prima dată de Corrado Gini cu 30 de ani înaintea teoriei informației – entropia lui Shannon, într-un studiu din *Atti del R. Inst. Veneta di Scienze, Lettere ed Arti*, 1917, 1918, v.LXXVII.

E de remarcat că ori de câte ori uniformitatea sau nedeterminarea sau dezorganizarea sistemului crește, energia sa informațională descrește. Așadar, energia informațională crește o dată cu diferențierea sau organizarea sistemului și descrește o dată cu uniformizarea sau dezorganizarea sistemului.

2. Utilizarea metodei energiei informaționale în construcția scalelor informaționale

În cele ce urmează am utilizat, spre exemplificarea metodei, rezultatele obținute la cercetarea concretă având ca obiectiv studiul fenomenului corupției, realizată prin chestionarul DAMIT și coordonată de dr. Dan Banciu (cercetare realizată în vara anului 1998).

Astfel, tabelul 1 prezintă distribuția opiniilor respondenților la întrebarea “Cât de grave sunt următoarele probleme sociale cu care vă confrunțați în prezent”. Sunt etalate astfel 10 tipuri de probleme sociale, grupate într-un bloc de probleme, cu posibilitatea ca răspunsurile să se poată distribui pe trei niveluri graduale: “grave”, “oarecum grave” și “deloc grave”.

Considerând fiecare din cele 10 probleme menționate ca fiind un sistem cu trei stări, vom putea calcula energia informațională a fiecărui sistem în parte. Ordonarea numerică a energiei informaționale pe cele 10 sisteme realizează o ierarhie standard. De remarcat faptul că în cazul **tabelului nr.1**, ierarhia

realizată de scala informațională “coincide” cu ierarhia oferită de scala graduală 1 (mai puțin locurile 9 și 10 care sunt schimbate între ele).

Tabelul nr.1
Cât de grave sunt următoarele probleme sociale cu care
vă confrunțați în prezent?

		Grave	Oarecum grave	Deloc grave	NS/NR	Energia informațională	SG ₁	SG ₂	SG ₃	S I
		1	2	3	4	0	1	2	3	
1	Nivel de trai	62.3	28.2	9.3	0.2	0.4763	4	3	6	4
2	Locuință	42.5	26.8	28.8	2.0	0.3357	7	6	4	7
3	Șomaj	41.1	25.4	29.4	4.2	0.3216	8	7	3	8
4	Protecție socială	50.0	28.0	17.3	4.8	0.3606	6	4	5	6
5	Criminalitate	68.5	19.6	7.90	4.0	0.5154	2	9	8	2
6	Corupție	80.2	11.9	5.20	2.8	0.6608	1	10	10	1
7	Reformă	53.6	36.3	5.6	4.6	0.4243	5	1	9	5
8	Mișcări sociale	17.9	35.7	37.9	8.5	0.3103	10	2	1	9
9	Venituri	65.1	25.0	8.7	1.2	0.4940	3	8	7	3
10	Alegeri anticipate	23.4	27.4	32.5	16.7	0.2633	9	5	2	10

Analiza informațională a **tabelului nr.2**, în care se analizează distribuția opiniilor subiecților la întrebarea: “*În ce măsură actuala situație socială din România se datorează următorilor factori?*” (și care sunt enumerați pe patru grade de intensitate), conduce la concluzia că scala informațională creează o ierarhie identică cu cea realizată de scala graduală 1 (“foarte mare măsură”).

Tabelul nr.2
În ce măsură actuala situație socială a României se datorează:

		În f. mare	În mare	În mică	În f. mică	Energie informativă	SG ₁	SG ₂	SG ₃	SI
		1	2	3	4	0	1	2	3	
1	Nerealizării reformei	84.5	11.7	1.2	2.6	0.7285	2	7	6	2
2	Evoluției prețurilor	80.8	14.5	2.6	2.2	0.6750	4	6	1	4
3	Moralității indivizilor	75.8	19.8	2.0	2.4	0.6147	8	1	2	8
4	Inflației	81.0	15.3	1.2	2.6	0.6803	3	5	7	3
5	Nerealizării privatizării	78.4	16.7	1.4	3.6	0.6440	5	4	4	5
6	Incompetențele guvernamentale	76.0	19.6	1.8	2.6	0.6170	7	2	3	7
7	Corupției	93.8	5.4	0.0	0.8	0.8828	1	8	8	1
8	Lipsei de autoritate a instituțiilor statului	77.0	17.1	1.4	4.6	0.6244	6	3	5	6

Nu același lucru se întâmplă în cazul **tabelului nr.3**, în care se analizează gradul de autoritate al unor instituții (13 la număr, pe trei grade de intensitate). Ierarhia realizată de scala informațională diferă esențial de toate scalele graduale (în număr de trei), cu păstrarea pozițiilor 1, 9 și 10 față de scala graduală 1.

Tabelul nr.3
Câtă autoritate (legitimare) credeți că au următoarele instituții?

		Mare/F. mare	Mică/F. mică	Deloc	NS/NR	Energie informațională	SG ₁	SG ₂	SG ₃	SI
		1	2	3	4	0	1	2	3	
1	Președinte	46.0	43.5	8.7	1.8	0.4097	3	9	12	2
2	Parlament	42.9	42.9	11.3	3.0	0.3817	4	10	11	7
3	Guvern	50.0	39.5	7.7	2.8	0.4127	1	11	13	1
4	Justiție	37.5	48.0	12.3	2.2	0.3866	6	5	9	5
5	Poliție	32.1	52.6	12.9	2.4	0.3969	7	3	8	3
6	Armată	49.0	35.1	11.3	4.6	0.3791	2	13	10	8
7	Garda financiară	31.7	47.2	14.3	6.7	0.3482	8	7	7	11
8	Dir. Vămii	24.6	46.2	18.7	10.5	0.3199	11	8	3	13
9	Administrație publică	23.0	55.0	15.9	6.2	0.3849	12	2	6	6
10	Partide politice	26.2	49.4	19.2	5.2	0.3522	10	4	2	10
11	Sindicate	19.4	56.5	18.1	6.0	0.3932	13	1	5	4
12	Biserică	42.5	36.3	18.3	4.0	0.3403	5	12	4	12
13	Serv. medicale	28.8	47.8	20.2	3.2	0.3532	9	6	1	9

Deosebit de relevant este cazul **tabelului nr.4**, în care se analizează *“În ce măsură s-au dovedit eficiente o serie de instituții în lupta împotriva corupției”*. Dacă în situațiile anterioare se impunea comparația ierarhiilor din scala informațională cu scala graduală 1 care este o scală extremală (de intensitate maximă), de această dată termenul de comparație îl oferă scala graduală de ordinul 2 (*“în mică măsură”*), în acest caz cele două ierarhii coincid.

Tabelul nr.4
În ce măsură s-au dovedit eficiente următoarele instituții în lupta împotriva corupției?

		Mare/F. mare	Mică/F. mică	Deloc	NS/NR	Energie informațio nală	SG ₁	SG ₂	SG ₃	SI
		1	2	3	4	0	1	2	3	
1	Poliția	22.8	61.3	12.7	3.2	0.4449	1	2	7	2
2	Justiția	19.8	62.9	13.3	4.0	0.4541	4	1	5	1
3	Procuratura	22.2	58.3	12.3	7.1	0.4093	3	5	8	5
4	Garda financiară	22.4	58.1	12.9	6.5	0.4086	2	6	6	6
5	Guvernul	18.8	59.3	17.9	4.0	0.4206	5	4	4	4
6	Primăriile	9.9	60.9	23.0	6.2	0.4374	6	3	3	3
7	Prefecturile	7.3	55.8	24.2	12.7	0.3913	8	7	2	7
8	Corpul gardienilor publici	7.9	39.7	39.5	12.9	0.3365	7	8	1	8

Nefiind apanajul scalelor graduale maximale, scala informațională se dovedește a fi o scală standardizată cu caracter de etalon pentru scalele intermediare. În același timp, ea poate oferi posibilitatea de a măsura abaterile scalelor graduale față de scala etalon.

Totodată, ea pune în evidență în mod pregnant scala care guvernează întregul complex de sisteme și, prin clasificarea standardizată care o creează, stabilește o ***ierarhie atitudinală***.

Capitolul XII

SISTEME ELECTORALE

1. Votarea prin electori

La un congres internațional de Matematică s-a propus să se analizeze următoarea situație – problemă: „În țara X, condusă de președintele Y se apropie alegerile prezidențiale. Țara X are 20 de milioane de alegători, dintre care numai un procent (1%) îl susține pe președintele Y, care urmează să organizeze alegerile în așa fel încât să fie reale și în același timp, alegerile să pară democratice. Este posibil acest lucru și cum anume?”

În mod curios, răspunsul este afirmativ. Într-adevăr, președintele Y urmează să împartă alegătorii în grupe egale. Acestea se vor împărți în continuare în subgrupe, ș.a.m.d. În subgrupele cele mai mici se alege un reprezentant – electorul. Electorii aleg reprezentanții lor pentru votare în grupele superioare, etc. În fine, electorii – reprezentanți ai grupelor celor mai mari aleg președintele. Să vedem cum poate fi pusă în practică această soluție, care ar da câștig la vot partidei minoritare. (Să ne amintim numai interesanta situație creată în toamna anului 2000 la alegerile prezidențiale din Statele Unite

ale Americii, în care un candidat a ieșit victorios ca urmare a votului majoritar al electorilor, deși după numărul de voturi al alegătorilor ar fi trebuit să câștige contracandidatul său).

Pentru înțelegerea schemei logice a soluției adoptate, să luăm situația din figura alăturată:

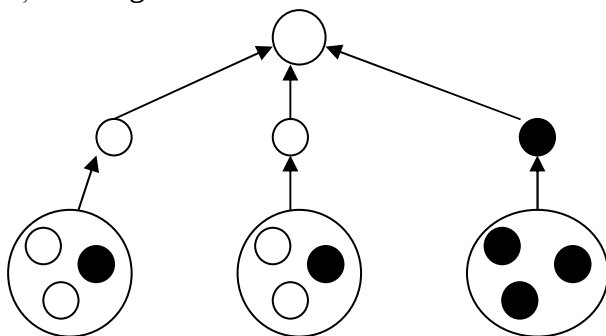


Fig. 1

După cum se poate constata, 9 alegători urmează să-și desemneze un lider. 4 sunt din grupa A și 5 din grupa B – majoritară. Grupările de pe nivelul 3, după cum se poate observa din figură, promovează trei electori, doi din grupa minoritară A și 1 din grupa majoritară B. Rezultatul votului final este evident: grupa minoritară își promovează liderul său!

Odată lămurit aspectul de principiu, să vedem cum se materializează soluția la problema inițială. Vom împărți cele 20 de milioane de alegători în 5 grupe de câte 4 milioane fiecare, astfel încât două dintre grupe să fie formate numai din adversari ai președintelui Y. Să numim aceste grupe de tip B, în

timp ce următoarele trei grupe de câte 4 milioane de alegători le vom numi de tip A. Fiecare dintre aceste grupe de rang 1 va fi împărțită în 5 subgrupe de rang 2, de câte 800.000 alegători, astfel încât din cele 5 grupe de rangul doi care formează o grupă de tip A de rangul întâi, trei să fie de tip A, etc. (v. fig. 2)

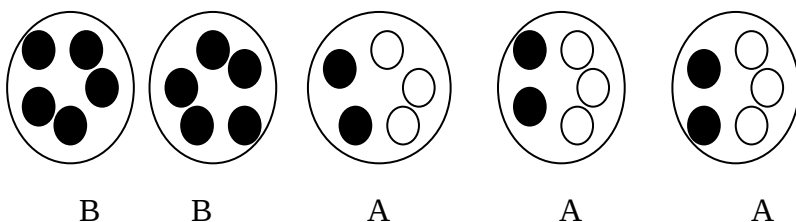


Fig.2

Procedeul se continuă după următoarea schemă logică: Fiecare grupă de rang r va fi împărțită la rândul ei în 5 subgrupe egale de rangul $r+1$, așa încât trei din cele 5 subgrupe de rangul $r+1$, care formează o grupă de tip A, de rangul r , să fie tot de tipul A. Situația finală se va prezenta ca în fig. 3:

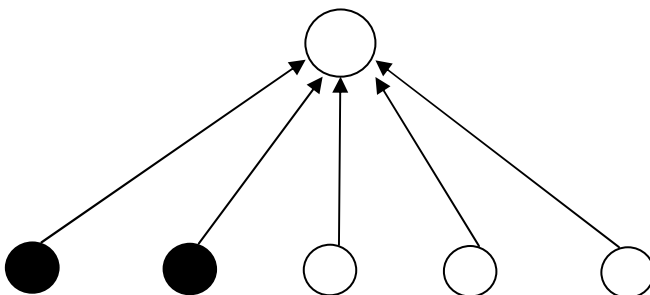


Fig. 3

care va da câștig de cauză partidei președintelui Y. Tabloul de mai jos este de fapt o schemă iterativă care ne arată după câți pași trebuie să ne oprim așa încât problema să capete răspuns afirmativ.

Continuând procesul de descompunere a fiecărei grupe în 5 subgrupe, dintre care 3 sunt de tip A și 2 de tip B se poate observa că la pasul 7 populația electorală va fi grupată în $5^7 = 78.125$ subgrupe, fiecare subgrupă conținând $2^8 = 256$ alegători. Dintre aceste subgrupe, $3^7 = 2187$ vor fi subgrupe de tip A, favorabile președintelui, iar restul de tip B. Dacă ne-am opri în acest moment, președintelui Y îi vor fi necesari $3^7 \cdot 256 = 2187 \cdot 256 = 559.872$ alegători pentru a câștiga alegerile. Cum el nu dispune de atâtea, va trebui continuat procesul de subgrupare. Dar o subgrupă de 256 alegători nu mai poate fi împărțită în cinci subgrupe egale. Așadar, grupa de 256 alegători va fi divizată în 16 grupe de câte 16 alegători, dintre acestea 9 fiind de tip A. În acest stadiu, președintelui Y îi sunt necesari $3^9 \cdot 16 = 315.928$ alegători favorabili, încă mai mulți decât dispune. O ultimă împărțire a grupelor, care va subdivide grupa de 16 alegători în 16 subgrupe, fiecare subgrupă, fiind formată de fapt din câte un alegător, iar dintre acestea, 9 vor fi de tip A, va conduce la rezultatul dorit. Într-adevăr, în acest moment mulținea alegătorilor este divizată în cele mai mici

subgrupe posibile, în total $2^8 \cdot 5^7 = 20.000.000$ subgrupe, fiecare subgrupă fiind constituită dintr-un singur alegător. 16 astfel de subgrupe, dintre care 9 sunt de tip A, vor forma cele $2^4 \cdot 5^7 = 1.250.000$ grupe de ordin imediat superior. Și tot așa, procesul se derulează după acest mecanism.

Așadar, după 9 pași, se obține victoria președintelui Y cu $3^{11} = 177.147$ alegători, care reprezintă mai puțin de 1% din totalul alegătorilor.

Iată că sistemul electoral al votului prin reprezentanții grupelor - electorii, poate conduce la surprize de proporții!

Nr. de pași	Numărul total al grupelor	Numărul grupelor de tip A	Numărul de alegători într-o grupă de rang r	Numărul de grupe de rang r+1 care constituie o grupă de rang r	Numărul subgrupe lor de tip A
Pasul 1	$5^1 = 5$	$3^1 = 3$	4.000.000	5	3
Pasul 2	$5^2 = 25$	$3^2 = 9$	800.000	5	3
Pasul 3	$5^3 = 125$	$3^3 = 27$	160.000	5	3
Pasul 4	$5^4 = 625$	$3^4 = 81$	32.000	5	3
Pasul 5	$5^5 = 3125$	$3^5 = 243$	6.400	5	3
Pasul 6	$5^6 = 15.625$	$3^6 = 729$	1.280	5	3
Pasul 7	$5^7 = 78.125$	$3^7 = 2187$	256	16	9
Pasul 8	$5^5 \cdot 2^4 = 16 \cdot 78.125$	$3^9 = 19.683$	16	16	9
Pasul 9	$2^8 \cdot 5^7 = 20.000.000$	$3^{11} = 177.147$	1	-	-

2. Principiul majorității

O dată cu apariția primelor constituții democratice, devenite realitate după victoria revoluției franceze de la 1789, s-a pus în discuție în mod firesc sistemul electoral al majorității. Nu mică a fost însă mirarea, când marchizul de Condorcet a analizat această regulă a majorității și a ajuns la concluzii paradoxale. Să urmărim îndeaproape raționamentul marchizului de Condorcet pe un exemplu. Să presupunem că 2100 de alegători votează pentru alegerea unui lider din 3 candidați desemnați, pe care îi vom nota prin A, B și C. Rezultatul votului arată după cum urmează:

	I	II	III
200 alegători preferă ierarhia	A	B	C
610 alegători preferă ierarhia	A	C	B
190 alegători preferă ierarhia	B	A	C
480 alegători preferă ierarhia	B	C	A
210 alegători preferă ierarhia	C	A	B
410 alegători preferă ierarhia	C	B	A

O analiză sumară ne arată că 810 alegători preferă candidatul A, 670 îl preferă pe B și 620 în preferă pe C. După regula majorității, pe primul loc se clasează A pe al doilea B și pe al treilea C, această ordine dându-i câștig de cauză candidatului A. Numai că la o analiză mai atentă, vom constata că A este preferat lui B de 1020 de alegători în timp ce B este preferat lui A de 1080 de alegători. Acest lucru spune că B

trebuie situat înaintea lui A în ordinea preferințelor electoratului. Mai mult, continuând analiza, vom constata că A este preferat lui C de 1000 de alegători, în același timp însă C este preferat lui A de 1100 persoane, adică C este situat și înaintea lui A, care rămâne astfel pe locul III. Pentru a vedea care este situația pentru locurile I și II, să mergem mai departe și să constatăm că B este preferat lui C de 870 de persoane în timp ce 1230 de alegători îl preferă pe C înaintea lui B. Această analiză stabilește următoarea ierarhie: C pe primul loc, B pe locul 2 și A pe locul 3. Acest rezultat, care intră în contradicție cu principiul simplei majorități, este cunoscut sub denumirea de **paradoxul lui Condorcet**.

Situația paradoxală generată de principiul majorității, face parte dintr-un context mai general, și anume, conduce la așa-numitele „alegeri multicriteriale”. În cazul nostru, candidații sunt obiecte de clasificat, iar alegătorii constituie criterii de clasificare (sau ierarhizare).

Există și alte numeroase exemple de acest fel. Deciziile de grup, deciziile de natură socială sau economică sunt în general multicriteriale, în care fiecare criteriu generează o ierarhie.

Din acest motiv s-au încercat numeroase metode de agregare a ierarhiilor, fără rezultate semnificative. Toate sunt

criticabile. Aplicând metode diferite aceleiași situații se obțin ierarhii diferite. Și atunci, s-a pus problema existenței sau non-existenței unei metode de sintetizare a ierarhiilor care să nu mai ducă la contradicții.

Răspunsul la această problemă l-a dat Keneth J. Arrow, profesor la Universitatea Harvard, și este din păcate negativ. Astfel, el a demonstrat că **nu există nici o metodă rațională de agregare a ierarhiilor.**

Conceptul de „raționalitate” a unei metode este definit de Arrow prin intermediul a cinci condiții firești, pentru orice metodă de agregare, și pe care le prezintăm după [2];

C₁. Există cel puțin trei obiecte de ierarhizat și cel puțin două criterii de ierarhizare (netrivialitatea problemei);

C₂. Dacă obiectul A este înaintea obiectului B în toate ordonările inițiale, atunci A este înaintea lui B și în ordonarea de sinteză (optimalitatea Pareto);

C₃. Ordinea finală a oricăror două obiecte depinde numai de ordinea lor în ierarhiile inițiale (condiția de independență);

C₄. Ordinea finală a oricăror două obiecte nu este impusă dinainte (condiția de suveranitate);

C₅. Ierarhia finală nu trebuie să fie identică cu o ierarhie inițial fixată (condiția absenței unei dictaturi).

Deși simple și aparent „inofensive”, se dovedește că mulțimea celor cinci condiții se constituie într-un corp contradictoriu. Altfel spus, îndeplinirea de către o metodă de agregare a 4 din cele cinci condiții, o va încălca pe a cincea. Așa de exemplu, îndeplinirea de către o metodă de agregare a ierarhiilor, a condițiilor $C_1+C_2+C_3+C_4$ va însemna că metoda respectivă aparține unui dictator.

Singurele posibilități pe care le oferă teorema lui Arrow sunt „relaxările”, uneia sau unor condiții sau renunțarea la una din cele 5 condiții.

Relaxarea condiției C_1 , în care rămân doar două obiecte de ierarhizat este și ea o soluție care nu antrenează paradoxul lui Condorcet. Poate și din acest motiv alegerile prezidențiale din al doilea tur de scrutin se fac numai pe doi candidați, cei mai bine plasați în cursa electorală.

Capitolul XIII

MODELE DE SUBIECTE

1. Cand adunam doua variabile aleatoare, probabilitatile evenimentelor elementare:
 - a) se aduna
 - b) se scad
 - c) se inmultesc
2. Cand inmultim doua variabile aleatoare, probabilitatile evenimentelor elementare:
 - a) se aduna
 - b) se scad
 - c) se inmultesc
3. Media sumei a doua variabile aleatoare este egala cu :
 - a) suma mediilor variabilelor aleatoare
 - b) produsul mediilor variabilelor aleatoare
 - c) diferenta mediilor variabilelor aleatoare
4. Media unei variabile aleatoare este:
 - a) egala cu cea mai mica valoare a variabilei aleatoare
 - b) egala cu cea mai mare valoare a variabilei aleatoare
 - c) cuprinsa intre acestea
5. Media produsului a doua variabile aleatoare este:
 - a) egala cu suma mediilor celor doua variabile aleatoare
 - b) egala cu produsul mediilor celor doua variabile aleatoare
 - c) egala cu radical din suma patratelor mediilor
 - d) nici unul din punctele a) b) c) nu este adevarat

6. Statistica descriptiva se ocupa cu :
- a) prelucrarea datelor statistice
 - b) culegerea datelor despre un fenomen social
 - c) inregistrarea datelor despre un fenomen social (b+c)
 - d) nici unul din raspunsurile a) b) c) nu este adevarat
7. Variabilele X si Y sunt independente daca :
- a) pentru o pereche de indici (i,j) evenimentele elementare corespunzatoare sunt independente
 - b) pentru mai multe perechi de indici (i,j) evenimentele elementare corespunzatoare sunt independente
 - c) pentru toate perechile (i,j) evenimentele elementare corespunzatoare sunt independente
8. Momentul de ordinal 2 al variabilei aleatoare X este:
- a) patraturul valorii medii a variabilei aleatoare X
 - b) valoarea medie a patraturului variabilei aleatoare X
 - c) alta expresie
9. Variabila aleatoare $X - M(X)$ se numeste :
- a) abaterea medie a variabilei aleatoare X
 - b) abaterea de la medie a variabilei aleatoare X
 - c) abaterea medie patratica a variabilei aleatoare X
10. Abaterea medie caracterizeaza :
- a) imprastierea variabilei aleatoare X
 - b) boltirea variabilei aleatoare X
 - c) concentrarea variabilei aleatoare X
11. Momentul centrat de ordinal II al variabilei aleatoare X este:
- a) abaterea medie a variabilei aleatoare X
 - b) abaterea de la medie a variabilei aleatoare X
 - c) dispersia variabilei aleatoare X
 - d) abaterea medie patratica a variabilei aleatoare X

12. Dispersia unei constante este egala cu :
- a) 0
 - b) 1
 - c) alta constanta
13. Dispersia sumei dintre o constanta si variabila aleatoare X este :
- a) acea constanta
 - b) egala cu 0
 - c) egala cu dispersia constantei
 - d) egala cu dispersia variabilei aleatoare X
14. Dispersia produsului dintre o constanta si variabila aleatoare X este egala cu :
- a) acea constanta
 - b) dispersia lui X
 - c) a $D(X)$
 - d) nici unul din raspunsurile a) b) c) nu este adevarat
15. Dispersia variabilei aleatoare X este egala cu :
- a) patratul abaterii medii patratic
 - b) radical din abaterea medie patratica
 - c) raspunsurile a) si b) nu sunt adevarate
16. Coeficientul de covarianta a doua variabile aleatoare este egal cu :
- a) media produsului variabilelor minus produsul mediilor variabilelor
 - b) media sumei variabilelor
 - c) suma patratelor mediilor variabilelor aleatoare
17. Amplitudinea unei variabile aleatoare este:
- a) un indicator de dispersie
 - b) un indicator al tendintei centrale de grupare
 - c) ambele la un loc
 - d) nici una din situatiile de mai sus

8. Amplitudinea unei variabile aleatoare este egala cu :
- a) suma dintre valoarea maxima si valoarea minima a variabilei aleatoare
 - b) media aritmetica a valorilor
 - c) diferenta dintre valoarea maxima si valoarea minima
19. Cuantilele de ordinal 10 ale unei variabile aleatoare se numesc:
- a) cuartile
 - b) centile
 - c) decile
20. Cuartila este o cuantila de ordinul :
- a) cinci
 - b) zece
 - c) patru
 - d) o suta
21. Mediana unei variabile aleatoare este egala cu :
- a) prima cuartila
 - b) a doua cuartila
 - c) a treia cuartila
22. Coeficientul de variatie este un indicator :
- a) al formei distributiei
 - b) de dispersie
 - c) al tendintei centrale de grupare
23. Oblicitatea este un indicator :
- a) al formei distributiei
 - b) de dispersie
 - c) al tendintei centrale de grupare

24. Boltirea este un indicator :
- a) al tendintei centrale de grupare
 - b) al formei distributiei
 - c) de dispersie
25. Media variabilei abaterea de la medie este egala cu :
- a) 0
 - b) 1
 - c) 0,5
26. Media produsului a 2 variabile aleatoare este egala cu :
- a) Suma mediilor v.a.
 - b) Produsul mediilor var.aleat.
 - c) punctul b) are loc cu conditia ca variabilele aleatoare sa fie independente.
27. Scalele nominale permit :
- a) O masurare a distantei dintre termenii scalei
 - b) O ierarhizare a valorilor
 - c) O enumerare de posibilitati.
28. Scalele ordinale permit :
- a) O ierarhizare a valorilor
 - b) posibilitatea unui raport intre 2 pozitii pe o scala
 - c) O masurare a distantei dintre termenii scalei
29. Scalele proportionale :
- a) permit o ierarhizare a valorilor
 - b) exprima posibilitatea unui raport intre 2 pozitii pe o scala
 - c) permit o enumerare de posibilitati.
30. Scalele cu intervale permit :
- a) Masurarea distantelor dintre diferitele trepte
 - b) O enumerare de posibilitati
 - c) Posibilitatea unui raport intre 2 pozitii pe o scala.

31. O variabila aleatoare se numeste discreta atunci cand :
- a) numerele care corespund modalitatilor ei sunt intregi
 - b) numerele care corespund modalitatilor ei sunt rationale
 - c) numerele care corespund modalitatilor ei sunt reale
32. Grupele de varsta constituie un exemplu de :
- a) scala de intervale
 - b) scala nominala
 - c) scala ordinala
33. Scalele care caracterizeaza variabilele ale caror modalitati nu au decat un scop descriptiv sunt :
- a) scale ordinale
 - b) scale nominale
 - c) scale de intervale
34. Raspunsurile de tip : « deloc/putin/mult/foarte mult », dintr-un chestionar, reprezinta un exemplu de scala :
- a) de raport
 - b) ordinala
 - c) nominala
 - d) de intervale
35. Scalele al caror statut se bazeaza pe faptul ca intervalele care separa doua modalitati succesive sunt intotdeauna egale intre ele sunt :
- a) scale ordinale
 - b) scale nominale
 - c) scale de intervale
36. Intr-un chestionar, intrebarile de opinie vizeaza :
- a) preocuparile intelectuale ale indivizilor
 - b) aspecte de comportament ale indivizilor
 - c) parerile, atitudinile subiectilor

37. Intr-un chestionar, intrebarile factuale privesc :
- a) aspectele de comportament ale subiectilor
 - b) parerile, latitudinile
 - c) preocuparile intelectuale
38. Esantionarea care are la baza principiul loteriei este :
- a) esantionare prin stratificare
 - b) esantionare multistadiala
 - c) esantionare simplu aleatoare
39. Esantionarea care presupune o grupare a populatiei pe arii geografice, culturale etc. este o esantionare :
- a) multifazica
 - b) prin stratificare
 - c) multistadiara (grupala)
40. Esantionarea care se efectueaza in populatii neomogene alcatuite din straturi, este o esantionare :
- a) simplu aleatoare
 - b) multistadiala
 - c) multifazica
 - d) prin stratificare
41. Esantionarea care presupune gruparea populatiei dupa cateva caracteristici, apoi determinarea subesantioanelor, este o esantionare :
- a) panel
 - b) pe cote
 - c) simplu aleatoare
 - d) prin stratificare
42. Erorile sistematice sunt erori de :
- a) reprezentativitate
 - b) de inregistrare
 - c) altele

43. Erorile intamplatoare sunt erori de :
- a) inregistrare
 - b) reprezentativitate
 - c) altele
44. Universul de esantionare este format din :
- a) populatia intrata in esantion
 - b) populatia intrata in studiu
 - c) toata populatia tarii
45. Procesul statistic prin care concluziile obtinute din studiul esantionului se pot extinde la intreaga populatie studiata se numeste :
- a) reprezentativitate
 - b) inferenta
 - c) analiza statistica
46. O ipoteza a modelului recursiv este ca :
- a) relatiile dintre variabile sa fie neliniare
 - b) relatiile dintre variabile sa fie liniare
 - c) sa nu existe astfel de relatii
47. Indicatorul concordantei pozitive ajuta la construirea :
- a) coeficientului de corelatie a doua variabile aleatoare
 - b) coeficientului de regresie
 - c) coeficientului de corelatie a rangurilor al lui Kendall
48. Inre indicatorul concordantei negative si coeficientul lui Kendall exista urmatoarea legatura :
- a) cresc in acelasi timp
 - b) unul creste in timp ce celalalt descreste
 - c) descrec in acelasi timp

49. Indicele lui Gini este un indicator :

- a) de dispersie
- b) al formei distributiei
- c) al tendintei centrale de grupare

50. In modelele explicative gradul de variabilitate al unui fenomen poate fi masurat de :

- a) indicatorii de dispersie
- b) indicatorii formei distributiei
- c) indicatorii tendintei centrale de grupare

Testul nr.1

- 1) Suma frecventelor relative ale unei serii de date este:
a) 1; b) 50; c) 100%
- 2) Variabila abatere medie caracterizeaza :
a) imprastierea; b) concentrarea ;c) forma distributiei.
- 3) Daca X si Y sunt 2 variabile aleatoare,atunci:
$$D(X+Y) = D(X) + D(Y)$$

a) intotdeauna; b) cand X si Y sunt independente; c)
Cand X si Y sunt dependente
- 4) Coeficientul de corelatie $r(X,Y)$ a doua variabile aleatoare este egal cu zero cand:
a) X si Y sunt independente; b) X si Y sunt dependente
- 5) Coeficientul lui Kendall masoara :
a) corelatia a doua variabile cantitative
b) corelatia rangurilor a doua v.a. oarecare
c) corelatia rangurilor a 2 v.a. calitative
- 6) Gradul de reprezentativitate al unui esantion depinde numai de:
a) caracteristicile populatiei
b) marimea esantionului
c) procedura de esantionare folosita
d) oricare din ele
e) de toate trei
- 7) Tipul de esantionare probabilist are caracter:
a) aleator
b) nealeator
c) oricare din ele
d) niciunul dintre acestea

- 8) Testul „hi patrat” reprezinta:
- a) un test de cunoastere
 - b) un test de semnificatie
 - c) un test de reprezentativitate
- 9) Modelul lui Blalock reprezinta:
- a) un model de analiza cauzala;
 - b) un model de analiza de dependenta
 - c) un model de comportament
- 10) Abaterea intercuartila este o caracteristica:
- a) standardizata
 - b) nestandardizata

Testul nr. 2

- 1) Frecventele relative sunt:
- a) numere intregi;
 - b) valori negative;
 - c) valori pozitive subunitare
- 2) Doua variabile aleatoare X si Y sunt independente daca:
- a) doua evenimente elementare sunt independente;
 - b) anumite evenimente elementare sunt independente;
 - c) toate evenimentele elementare sunt independente doua cate doua.
- 3) Cuartila este un indicator care caracterizeaza :
- a) tendinta centrala de grupare;
 - b) imprastierea variabilei;
 - c) forma distributiei unei v.a.
- 4) Corelatia rangurilor a 2 v.a. se exprima cu ajutorul
- a) coeficientului lui Kendall;
 - b) coeficientului lui Spearmann;
 - c) coeficientului lui Pearson

- 5) Modelul recursiv este un model de:
- a) analiza de dependenta;
 - b) analiza cauzala;
 - c) analiza de regresie.
- 6) Gradul de reprezentativitate al unui esantion este masurat de:
- a) coeficientul de corelatie;
 - b) coeficientul de varianta;
 - c) eroarea maxima admisa;
 - d) nivelul de incredere.
- 7) testul „hi patrat” se aplica:
- a) esantioanelor mici;
 - b) esantioanelor mijlocii;
 - c) esantioanelor mari.
- 8) Non-raspunsuri intr-un chestionar de opinie inseamna:
- a) raspunsuri negative;
 - b) lipsa raspunsurilor la unele intrebari;
 - c) lipsa raspunsurilor la intregul chestionar.
- 9) Cate tipuri de scale intalnim in stiintele sociale:
- a) 2 tipuri;
 - b) 3tipuri
 - c) 4 tipuri.
- 10) Pentru analiza seriilor interdependente se foloseste care din urmatoarele metode statistice:
- a) regresia;
 - b) indicatorul concordantei pozitive;
 - c) corelatia.

Testul nr. 3

- 1) Coeficientul de variatie este un indicator:
 - a) de pozitie;
 - b) de imprastiere;
 - c) al formei distributiei v.a.

- 2) Abaterea intercuartila relativa este un indicator :
 - a) standardizat;
 - b) nestandardizat

- 3) Indicatorul de boltire masoara :
 - a) forma distributiei unei v.a.
 - b) oblicitatea distributiei;
 - c) dispersia v.a.

- 4) Care afirmatie este adevarata :
 - a) Daca variabilele X si Y sunt independente atunci coeficientul lor de corelatie este nul;
 - b) Daca coeficientul de corelatie a 2 v.a. este nul atunci variabilele X si Y sunt independente;
 - c) Amandoua sunt adevarate.

- 5) Intr-un chestionar de opinie,intrebarile factuale sunt intrebari care privesc:
 - a) continutul acestora;
 - b) forma de inregistrare a raspunsurilor.

- 6) O procedura de esantionare in care fiecare individ din populatie are o sansa reala,calculabila si nenula de a fi ales in esantion,se va numi:
 - a) aleatoare;
 - b) nealeatoare.

7) Daca coeficientul de corelatie a doua v.a. este 1 atunci intre cele doua variabile exista o dependenta:

- a) neliniara;
- b) liniara;
- c) logaritmica;

8) Erorile statistice in realizarea unui esantion apar:

- a) cand nu se respecta principiile de baza ale teoriei esantionarii;
- b) in timpul derularii sondajului in teren;

9) Esantionarea care presupune gruparea populatiei dupa cateva caracteristici este specifica:

- a) esantionarii panel;
- b) esantionarii multifazice;
- c) esantionarii pe cote;

10) In cazul acceptarii ipotezei nule care varianta este posibila:

- a) ipoteza facuta este adevarata si urmeaza sa fie acceptata;
- b) ipoteza facuta este falsa;

Testul 4

1) Indicele lui Gini este un indicator:

- a) de pozitie ;
- b) de imprastiere;
- c) al formei distributiei variabilei aleatoare.

2) Coeficientul de corelatie a doua variabile aleatoare este:

- a) standardizat si ia valori in intervalul $[-2;2]$;
- b) standardizat si ia valori in intervalul $[-1;1]$;
- c) nestandardizat si ia valori pozitive.

3) Afectarea caracterului aleator in realizarea unui esantion este o eroare:

- a) sistematica;
- b) intamplatoare;
- c) de alta natura.

4) Cresterea in volum a esantionului asigura intotdeauna cresterea gradului de reprezentativitate?

- a) da;
- b) nu;
- c) depinde de alti parametri..

5) Indicatorul concordantei pozitive influenteaza marimea coeficientului de corelatie a rangurilor a lui Kendall?

- a) da;
- b) nu;
- c) nu intotdeauna;

6) Intr-un model recursiv relatiile dintre variabile trebuie sa fie, prin ipoteza:

- a) liniare;
- b) neliniare;
- c) oricum;

7) O caracteristica se numeste continua atunci cand numarul valorilor este:

- a) finit;
- b) infinit;
- c) cand valorile sunt pozitive.

8) Daca $M(X)$ este media v.a. X atunci proprietatea $M(XY) = M(X)M(Y)$ are loc cand:

- a) X si Y sunt dependente;
- b) X si Y sunt independente;
- c) oricand.

9) Stabilirea gradelor de libertate este importanta in:

- a) utilizarea testului „hi patrat”;
- b) realizarea unui model recursiv
- c) aplicarea ipotezei nule.

10) Mediana este:

- a) prima cuartila;
- b) a doua cuartila;
- c) a treia cuartila.

Bibliografie selectivă

1. Johan Galtung, *theory and Methods of Social Research*
2. N. Mihăilă, *Introducere în teoria probabilităților și statistică matematică*
3. GH. Mihoc, V.Urseanu, *Matematici aplicate în statistică*
4. Dumitru Porojan, *Statistica și teoria Sondajului*, Ed. Șansa, 1993
5. T. Rotariu, P. Iluț, *Ancheta sociologică și sondajul de opinie*, Ed. Polirom, 1997
6. T. Rotariu (coord.) ș.a., *Metode statistice aplicate în științele sociale*, Ed. Polirom, 1999
7. Ioan Mărginean, *Măsurarea în sociologie*, Ed. Șt. și Enciclopedică, 1982
8. D. Sandu, *Statistică în științele sociale*, București, 1992
9. Yule, G.U., Kendall, M.G., *Introducere în teoria statisticii* Ed. Științifică, 1969
10. Moser C.A., *Metodele de anchetă în investigarea fenomenelor sociale*, Ed. Șt. (1967)
11. Durkheim, E., *Regulile metodei sociologice*, Ed. Șt., 1974
12. Chelcea S., *Chestionarul în investigația sociologică*, Ed. Șt. și Enciclopedică, 1975

13. Alexandrescu P., *Sistemele electorale. Principiul majorității și limitele sale*. Rev. Română de Sociologie, nr.5-6/2001
14. Alexandrescu P., *Modalități de ameliorare a cercetărilor de teren care au la bază sondajul statistic*, Rev. Română de Sociologie, nr.1-2/2000
15. Alexandrescu P., *O metodă de analiză scalară și ierarhizare*, Rev. Română de Sociologie, nr.3-4/2000